Harmful Speech

Daniel Pi*

September 18, 2018

Contents

1 The Argument for Harm-Based Regulation 3
  1.1 General Theories of Speech Rights . . . . . . . . . . . . . . . . . . . 7
  1.2 The Structure of First Amendment Review . . . . . . . . . . . . . . 12
  1.3 Defamation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 14
  1.4 Espionage Act Cases and Incitement . . . . . . . . . . . . . . . . . 18
  1.5 Fighting Words and Hate Speech . . . . . . . . . . . . . . . . . . . . 21
  1.6 Commercial Speech . . . . . . . . . . . . . . . . . . . . . . . . . . . 28
  1.7 Obscenity . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 31
  1.8 General Theories of Free Speech . . . . . . . . . . . . . . . . . . . . 38

2 Refutation of the Harm-Based Argument 39
  2.1 Endogenizing the Harm of “Harmful Speech” . . . . . . . . . . . . . 39
  2.2 Endogenizing the Benefit of Harmful Speech . . . . . . . . . . . . . 41
  2.3 Filtering . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 43
  2.4 Counter-Argument in General Form . . . . . . . . . . . . . . . . . . . 46
  2.5 Counter-Counterarguments Anticipated . . . . . . . . . . . . . . . . . 47

3 Applications 49
  3.1 Affronting Speech . . . . . . . . . . . . . . . . . . . . . . . . . . . . 50
  3.1.1 Offense . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 50
  3.1.2 Obscenity, Indecency, and Profanity . . . . . . . . . . . . . . . . . 53

*Visiting Assistant Professor, George Mason University, Antonin Scalia Law School. Contact: dpi@gmu.edu. Utmost thanks go to my ever-industrious and incalculably talented research assistants, Kevin Bell, Rakin Hamad, Tomasz Mielniczuk, and Alec Phillips.
The freedom of expression is a defining characteristic of liberal democracy.\textsuperscript{1} Yet in actual practice, exceptions are invariably allowed to pervade, which permit governments to regulate disfavored categories of speech. A predominant justification given for such regulation is that some classes of expressive activity are supposed on balance to be harmful. This article presents an argument undermining that justification. I contend that merely identifying a category of speech as generating low social benefit and high social cost is an insufficient ground to justify regulation of that speech.

The arguments I oppose are methodologically consequentialist and economic in nature. My counterargument is correspondingly formulated within that same analytical framework. Certainly not all arguments for speech regulations are of this kind. Alternatives exist. For example: that the law should be an instrument for preserving a community’s “cultural identity,” or that a class of expressive activity is “contrary to the will of God.” The counterarguments I present in this article are not aimed at addressing those positions. Though I am deeply skeptical of these alternative justifications, I am keen to keep my present investigation focused, and

\textsuperscript{1}Indeed, even the most brutally illiberal governments have at least nominally recognized a general speech right. For example, presumptions against the regulation of speech were enshrined in the constitutions of Nazi Germany, Weimar Constitution art. 118 (1919), the Soviet Union, Constitution of the Russian Socialist Federated Soviet Republic, art. 2.14 (1918), Maoist China, Constitution of the People’s Republic of China, art. 87 (1954), and North Korea, Constitution of North Korea, art. 67 (1972, rev. 1998). Although Libya under Gaddafi lacked a formal constitution, it was nevertheless the official position of the Libyan government that citizens should enjoy unlimited expressive rights. Gaddafi, Muammar. The Green Book (1975) (“An individual has the right to express himself or herself even if he or she behaves irrationally to demonstrate his or her insanity.”). Indeed, Libyan law extended full freedom of expression even to commercial speech, even specifying tobacco advertising as an example where speech rights trump public health interests. However, Gaddafi also proclaims, “private individuals should not be permitted to own any public means of publication or information,” id., which effectively rendered the right irrelevant. Moreover, even residual “non-public speech rights” tended to be disregarded in practice in Gaddafi’s Libya—as in the aforementioned authoritarian regimes.
therefore bracket off those issues as topics for possible future research.

My exposition consists in three main parts. In the first, I provide a generalized construction of the harm-based justification for speech regulation and establish its prevalence in jurisprudential and scholarly thought. Next, I identify three countervailing effects, which undermine the logic of the received view. In the presence of these countervailing effects, I demonstrate that government efforts to reduce harmful speech will be less effective than its proponents have supposed, and that they may even exacerbate the harms they were intended to remedy. Finally, I sketch out how the three effects may arise in specific speech contexts.

1 The Argument for Harm-Based Regulation

Let us begin by abstracting the class of speech restrictions which limit an individual’s right to expression in cases where it generates negligible social value and imposes a substantial negative externality. I take these two conditions to define “harmful speech,” demarcating the scope of the present inquiry.

Definition. “x is harmful speech” is true if and only if

Condition (i). x generates negligible social value, and

Condition (ii). x imposes a substantial negative externality.

The main result of this article requires weaker criteria than I have defined. The principles I expost are applicable whenever social welfare maximization is used to justify speech regulation. However, for practical purposes it is more useful to constrain the discussion in terms of conditions (i) and (ii), because courts and scholars have tended to formulate harmfulness in those terms.

Several general reasons are commonly given in opposition to speech regulations. First, that it is difficult to quantify and compare the value and social cost of speech. Second, that the risk of “slippery slope” effects abound. Third, that citizens of liberal democracies derive utility from the very possession of the legal right itself—beyond its merely instrumental usefulness. In light of these reasons, harm-based rationales for speech regulation usually assume that social costs of a class of expression are not merely greater than benefits, but substantially greater than its benefits. For this reason, the conditions are formulated in terms of “negligible” social value and “substantial” negative externality. The principles exposted follow a fortiori for all harm-based justifications for speech regulation.

Arguments for the limitation of speech rights assume, of course, the existence of a general speech right. In the United States, this is established in the First Amendment
of the Constitution. The extent of the general right is unspecified in the text, and it is generally understood as the prerogative of the judicial system to determine its boundaries. Broadly, the courts have adopted two approaches to limiting the speech right. First, the courts can deem a category of behavior to fall outside the ambit of speech. In other words, the courts can decline to legally recognize an activity as being speech, even though it has some incidental expressive component. And, being non-speech, it falls outside the protection of the First Amendment. The government might therefore regulate it—assuming of course that the proposed regulation otherwise falls within the powers which the Constitution assigns to the government. Second, the courts can acknowledge a category of behavior as being speech properly, but hold that public policy demands an exception. The most forceful public policy rationales

---

2 U.S. Const. amend. I.
3 There is a reasonable textualist argument that the extent is specified. “Congress shall make no law . . . abridging the freedom of speech, or of the press,” implies that any law abridging speech violates the rights of citizens. Id. Courts have declined to accept this view, and the “no law” language is customarily treated as something less than absolute.
4 The battle lines are somewhat less clear than I have indicated when viewed up close. For example, the “traditionally exclusions” of the general speech right are sometimes argued to be non-speech, and sometimes argued to be speech but within a recognized exception. These “traditional” categories include defamation, obscenity, and incitement. The distinction is important in the structure of the argument. If the categories are defined as falling outside speech, then no justification for their regulation is needed. However, if they fall within the ambit of “speech,” then their regulation does require justification. To the extent the latter view is held, the arguments I present in this article will apply.

The arguments given for regarding such expressions as non-speech are often “originalist” in nature. The originalist argument is that these suspect speech categories were assumed to be permissible in the English law, and that the framers of the Constitution would not therefore have regarded nor intended the First Amendment to be a bar on their regulation.

This is clearly mistaken. James Madison, addressing Congress on the adoption of the Bill of Rights, expressly sought a break from past practice, distinguishing the guarantees of individual liberty from earlier political and legal norms, saying, “But although . . . it may not be thought necessary to provide limits for the legislative power in [Britain], yet a different opinion prevails in the United States,” and “The right of freedom of speech is secured; the liberty of the press is expressly declared to be beyond the reach of this Government. . . .” 1 Annals of Cong. 436, 738 (1789).

Madison does not say, “unless of course the Government’s interest were compelling, in which case, infringe away!”

Yet lest I be seen to endorse amateurish historical speculation, I should add that I am skeptical what value there is in the originalist mode of Constitutional interpretation, that I do not pretend to be a historian, and that I question how seriously we should take the moral speculations of bewigged, pre-Industrial-Revolution plantation owners. I point to Madison’s remarks merely to indicate an “even if” argument. Regretfully, a fuller discussion of the topic would require a digression of such magnitude as to overwhelm my present thesis.
typically involve the identification of some social harm, which the regulation of speech is meant to remedy.

There is little difference between the two approaches in effect. Whether the courts declare a category of conduct to be “not speech,” or “speech for which an exception to the general right exists” is immaterial. The justifications for both approaches—to the extent that the justification is grounded in the harmfulness of a speech category—will employ essentially identical reasoning.

Suppose the government proposes to regulate a class of expressive activity. Call it $x$. Harm-based arguments advocating for the regulation of $x$ share the general form:

1. $x$ is harmful speech.

2. If harmful speech is punished, then social welfare is improved.

3. If $x$ is punished, then social welfare is improved. (*from 1, 2*)

4. If punishing $x$ improves social welfare, then the government should be permitted to punish $x$.

5. Therefore, the government should be permitted to punish $x$. (*from 3, 4*)

The target of my counterargument is premise (2). If it can be shown that (2) is false, then the inference to (3) fails. And without (3), the inference to (5) fails.

It bears remarking that premise (4), as I have formulated it, is also vulnerable to counterargument. Concededly, it could be nuanced to better reflect the varying levels of judicial scrutiny which the courts have applied to First Amendment controversies. Regardless, premise (4) is not the target of my present inquiry, and if I have stated (4) too extravagantly, this does no damage to my counterargument, for I stipulate to the point arguendo. Granting premise (4) only strengthens the argument I oppose.

Proponents of speech regulations do not assume premise (2) axiomatically. It is supportable with further argumentation. For convenience, let us call the supporting argument the “harm-deterrence argument”:

2.1. If harmful speech is punished, then the supply of harmful speech decreases.

2.2. If the supply of harmful speech decreases, then social welfare is improved.

2. If harmful speech is punished, then social welfare is improved. (*from 2.1, 2.2*)
This can of course be analyzed still further. I expect premise (2.1) to be founded on a model of deterrence, borrowed from the standard economic analysis of criminal law.\(^5\) And premise (2.2) seems to follow trivially from the definition of “harmful speech” (conditions (i) and (ii) above).

The harm-deterrence argument has a kind of elegance, benefitting from its resemblance to the economic approach to criminal and tort law.\(^6\) By association with that scholarship, it feels familiar, uncomplicated, and vaguely “right.”

Of course, the explicitly economic reconstruction of the argument for harm-based regulation is seldom stated in the general form I have articulated. It is however not entirely absent in the scholarly literature. Judge Richard Posner, ever the pioneer, attempted a general economic analysis of speech rights in *Free Speech in an Economic Perspective*. In that article, Posner devises a cost-benefit framework for interpreting the First Amendment, which is essentially equivalent to the harm-deterrence argument above.\(^7\) Unfortunately, Posner’s initial push, which he carefully distinguished as being “partial” and “tentative,”\(^8\) has not succeeded in inspiring many economists to join the effort. In the three decades since its publication, Posner’s article remains the most comprehensive treatment of speech rights from an overtly economic perspective.

It would however be a mistake to characterize the harm-deterrence argument as Posner’s argument. The policy rationales of non-economists used to justify speech regulation are implicitly economic arguments, which tacitly rely upon the harm-deterrence argument for their foundation. Indeed, most of the frequently rehearsed arguments for speech regulation are really harm-based economic justifications in disguise. This last assertion, of course, requires some defense.

To foreclose potential complaints that I am setting up a straw man, some examples of courts and legal scholars utilizing the harm-based justification for speech regulation are wanted. The remainder of this section addresses this concern.

In the interest of clarity, it is worth specifying precisely what I set out to do in subsections 1.2–1.8 below. I aim to show that the harm-based justification is a common thread running through free speech case law and scholarship. Moreover, that in the absence of a harm-based justification for regulation, the general right to free speech controls. In other words, subsections 1.2–1.8 are meant to establish that

\(^{5}\text{See Gary Becker, *Crime and Punishment*, 76 J. Pol. Econ. 169 (1968).}\)

\(^{6}\text{See, e.g., Becker, supra note 5; Steven Shavell, *Economic Analysis of Accident Law* (Harvard University Press 2007).}\)


\(^{8}\text{*Free Speech*, supra note 7, at 3.}\)
courts and legal scholars are committed to the proposition that the general speech right can be abridged if and only if there is a harm-based justification for carving out an exception. To demonstrate this, I provide a capsule tour of First Amendment jurisprudence, highlighting the harm-based argument underlying the justifications for regulation and non-regulation throughout history and across speech types.

For example, if a judge indicates that the low social value and high social costs of a class of expressive activity are grounds for allowing regulation of it, we may infer that he is assuming the harm-based argument diagrammed above. And if a judge points out either the non-negligible value or innocuousness of a class of expressive activity as a ground for protecting it, he must be assuming that: (1) the harm-based justification is the only relevant justification for abridging the general speech right; and (2) that the type of expressions at issue commonly fail to satisfy condition (i) or condition (ii).

In either example, the inferential connection between the relative harms and benefits of the speech category depends upon the harm-based justification. The reference to harms in the context of speech regulation would not make sense if the validity of harm-based justifications were not an assumed major premise.

1.1 General Theories of Speech Rights

Innumerable general theories of free speech have been offered throughout history. In this subsection, I seek merely to offer a handful of representative samples, highlighting the prevalence of harm-based justifications for allowing abridgment of the general speech right. Unavoidably, the task entails some insensitivity to the nuances of the authors’ positions. The reasons given for when and why speech rights should be respected, and when and why they should make way for other concerns are diverse and subtle. However, these subtleties are mostly irrelevant for present purposes. What matters is that the authors accept some specification of the harm-based argument for abridging speech rights.

The notion of a general speech right is not a recent invention. The ancient Athenians claimed to be the first to recognize a general freedom of speech, although, recalling the death of Socrates, we may well wonder as to its true extent. The modern origins of free speech might be located in John Milton’s Areopagitica, although, recalling the death of Socrates, we may well wonder as to its true extent. The modern origins of free speech might be located in John Milton’s Areopagitica, although, recalling the death of Socrates, we may well wonder as to its true extent.
ten praised as among the most compelling pleas for liberal governance. However, the reverence in which civil libertarians hold the poet is puzzling, for he seemingly would allow considerable incursion upon the speech right.\textsuperscript{11} The concern which motivated the \textit{Areopagitica} was the licensing and censorship of printed material, and it is only against these forms of speech regulation which Milton protested. Milton advocated no limitation whatever on how far governments should punish the authors of blasphemous publications after the fact.\textsuperscript{12} In the present-day parlance of law, he opposed merely the “prior restraint” of speech. He was apparently an enthusiastic supporter of ex post regulation. Yet even with respect to prior restraint, Milton’s commitment seems tepid relative to present-day standards. He was generous in the exceptions he would carve out of the prohibition on prior restraints of speech—conceding to restrictions on “popery,” superstition, impiety, or “evil” speech.\textsuperscript{13}

Certainly, Milton’s arguments—both for and against speech rights—were predominantly moralistic. Moreover, his arguments for liberty were narrowly focused on freedom from prior restraints. Nevertheless, the \textit{Areopagitica} is a landmark in the evolution of free speech, and it merits at least passing consideration whether even here the harm-based justification may be retrieved. Inasmuch as “popery” and “open superstition” are deemed sufficiently extreme to warrant the prior restraint, so too are expressions which have the tendency to extirpate “civil supremac[y].”\textsuperscript{14} This at least seems unambiguously to be an exclusion to Milton’s prohibition on prior restraint founded upon the reduction of secular harms.

Immanuel Kant is another thinker often characterized as an ardent defender of speech rights. Given his association with deontological ethics, it may surprise some readers to learn that he seems at least in some of his work to have endorsed harm-based justifications for abridging the right to free speech. For example, in \textit{What is Enlightenment?},\textsuperscript{15} he writes, “Nothing is required for this enlightenment, however, except freedom, and indeed the most harmless among all the things to which this term can properly be applied. It is the freedom to make public use of one’s reason at every

\textsuperscript{11}His arguments for speech rights are unimpressive also. Much of his polemic is aimed at establishing the value of speech for the Christian religion. Quite a lot of his argument is founded upon appeals to authority and vague references to the “will of God.”

\textsuperscript{12}CITE. “Those which otherwise come forth, if they be found mischievous and libellous, the fire and the executioner will be the timeliest and the most effectuall remedy, that mans prevention can use.”

\textsuperscript{13}CITE. “I mean not tolerated Popery, and open superstition, which as it extirpats all religions and civill supremacies, so it self should be extirpat, . . . that also which is impious or evil absolutely either against faith or maners no law can possibly permit, that intends not to unlawful it self.”

\textsuperscript{14}\textit{Id.}

\textsuperscript{15}CITE
point.” Kant’s gratuitous characterization of speech as “harmless” reveals much about his thinking. Why should it matter that the freedom of speech is “harmless,” unless harm were a possible ground for abridgment of the right? Indeed, Kant seems to have believed precisely this, for he goes on to write that the speech of public officials and pastors might justifiably be abridged on account of the harms which would result from unencumbered expression of their thoughts, owing to the authority of their offices. Harms, therefore, can justify the abridgment of speech, even in Kant’s relatively liberal view. It merely happens to be that in Kant’s view most speech simply cannot be very harmful.

John Stuart Mill adopts a similar position—though his underlying political and moral theory is quite distinct from Kant’s. Mill was arguably most important proponent of speech rights—certainly prior to the twentieth century.

The first two chapters of Mill’s On Liberty express in crystalline prose many fine arguments for a liberal speech doctrine. He is generally understood as advancing an extremum position, maximally favoring the right to free speech. Yet even Mill would allow abridgment of the speech right in order to mitigate harms. He is explicit on this point, writing, “[T]he only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others.” The principle was not merely a latent possibility for Mill. He furnished concrete examples of the abridgments he had in mind. In the third chapter, he writes:

> [E]ven opinions lose their immunity, when the circumstances in which they are expressed are such as to constitute their expression a positive instigation to some mischievous act. An opinion that corn-dealers are starvers of the poor, or that private property is robbery, ought to be unmolested when simply circulated through the press, but may justly incur punishment when delivered orally to an excited mob assembled before the house of a corn-dealer, or when handed about among the same mob in the form of a placard. Acts of whatever kind, which, without justifiable cause, do harm to others, may be, and in the more important cases absolutely require to be, controlled by the unfavourable sentiments, and, when needful, by the active interference of mankind. The liberty of the individual must be thus far limited; he must not make himself a

---

16 Id.

17 A Kantian basis for further abridgment of speech rights may be found, for example, in Helga Varden, *A Kantian Conception of Free Speech, in Free Speech in a Diverse World* 39 (Deidre Golash ed., 2010).

nuisance to other people.¹⁹

Mill’s position is clear: the prevention of harms can justify abridgment of the speech right. The example he supplies would fall within the category of “incitement” in present-day American jurisprudence, however the reasoning he employs seems easily extensible to other cognizable exceptions—for example, the publication of state secrets, true threats, and communications facilitating criminal conspiracies. It is unlikely however that Mill would allow much more than this.

Having opened the door to harm-based justifications, Mill’s reluctance to admit exceptions for offensive, obscenity, false or libelous statements, or otherwise harmful expressions may seem puzzling. On this point, it is worth clarifying that Mill’s definition of “harm” is different from mine. Mill’s conception is narrower. He does not believe that any harm can justify the abridgment of the speech right, because many of the things which satisfy conditions (i) and (ii) would not count as proper “harms” by his lights. He would exclude psychic harms—for example, shock, disgust, loss of dignity—as non-harms. For an expression to justify exception to the general speech right, its harm must be both substantial and physical. Mill would surely object to most of the speech regulation which liberal democracies presently allow. Yet he would not reject their harm-based justifications on principle. Rather, he seems to have believed that the proper role of governments to redress certain kinds of harm excluded by his definition.

Mill’s insensitivity to psychic harms is somewhat surprising. His utilitarian moral philosophy—equating “pleasure” with the good, and “pain” with bad—seems to ground all ethics upon wholly mental phenomena. He was moreover an economist, making it all the more puzzling why he refrains from drawing the natural equivalence between “harm” with disutility. Possibly he was concerned about the problem of error and the inaccessibility of mental states to third-party observers. Regardless, little interpretative work is required to see the harm-based justification at work in Mill’s theory of speech rights. Though not all harms—as defined by conditions (i) and (ii)—can justify speech regulation, any acceptable abridgment of the speech right, according to Mill, requires some harm-based justification.

Many subsequent moral and political philosophers have accepted some variation of Mill’s harm principle. For example, Jeremy Waldron advances a theory substantially similar to Mill’s, though expanding the scope of the definition of “harm” to include hate speech.²⁰ For Waldron, psychic harms are no less “harm” for the purposes of speech rights analysis than physical harms. More generally, philosophers,

¹⁹Id. at 54???.
psychologists, and the lawyers have tended increasingly to recognize psychic and dignitary harms as properly warranting legal redress, and the extension to Mill’s harm principle seems a natural and sensible evolution of the theory.

A far more ambitious expansion of Mill’s theory may be found in Professor Feinberg’s four-volume series, *The Moral Limits of the Criminal Law*. In the first volume, Feinberg articulates a definition of “harm,” which is distinct both from my definition and Mill’s definition: “[O]nly setbacks of interests that are wrongs, and wrongs that are setbacks to interest, are to count as harms in the appropriate sense.”21 Interests and wrongs are given careful treatment in succeeding sections. The relevant point is that, like Mill, Feinberg’s conception is constrained to exclude quite a lot of what we might ordinarily mean by “harm.” Unlike Mill, Feinberg’s aim was not to provide a general normative theory of law nor of justified government interference, but rather more narrowly—as his title indicates—of the criminal law. Nevertheless, quite a lot of speech regulation occurs in the form of criminalization, to which Feinberg’s theory is directly relevant. Moreover, to the extent that Feinberg’s principles are generalizable, it is indirectly informative to non-criminal speech regulations.

Feinberg’s harm principle includes things which I would bracket off as “moral.” He therefore would reject the strong claim that only expressions satisfying conditions (i) and (ii) can justify abridgment of the speech right. However, he does seem to believe that “harms” (as I have defined them) are one of the possible justifications for the abridgment of speech rights. He is therefore at least committed to the harm-based justification. Importantly, Feinberg’s definition of “harm” does not include the kinds of harms which Waldron would regard as falling within the harm principle—i.e., psychic harms.

Yet the exclusion of psychic harms from the definition of “harm” does not entail that criminal punishment on the basis of psychological effects cannot be justified under Feinberg’s theory. The harm principle is not the only operative justification in his framework. Feinberg would also allow abridgment of speech rights on the basis of “offense,” the topic of the second volume in his series.22 Not all offense can justify a criminal sanction. Similar to his narrowing of the meaning of “harm,” Feinberg adopts a notion of “serious offense” which requires an element of wrongfulness. For present purposes, the distinction Feinberg draws between “harms” and “offenses,” is immaterial. Expressions which cause “harm” or “offense” (by Feinberg’s definition) are both “harmful speech” (under my definition).

1.2 The Structure of First Amendment Review

Some remarks about the structure of First Amendment constitutional review may be helpful in organizing the discussion that follows. First Amendment review developed alongside mid-century innovations in Equal Protection and Due Process jurisprudence, recognizing multiple “levels” of constitutional scrutiny. Tiered review was first established in *United States v. Carolene Products Co.* Articulation of the “strict scrutiny” standard followed in the cases shortly thereafter. These seminal opinions averred that core freedoms—which include the freedom of speech and freedom of the press—would trigger the most stringent standard of review when implicated. And as litigation of First Amendment claims grew more frequent, the courts found it expedient to capitalize upon the opportunity to extend the doctrine to First Amendment disputes.

The nutshell gloss of First Amendment judicial review is that “content-based” regulations warrant strict scrutiny, and “content-neutral” regulations intermediate scrutiny. The rationale is that incursions upon the ideas and beliefs of the citizenry are anathema to the principles of a liberal democracy. Content-based restrictions target semantic content—the very thing feared of an oppressive government. Whereas content-neutral restrictions target conduct, which only incidentally affects the marketplace of ideas, so long as alternative channels of communicating the same information remain available to speakers. It is the distinction between regulating what is said versus how, when, and where it is said. Thus, if the speech affected by a content-neutral regulation can be expressed via alternative channels, the expected damage to the free exchange of ideas would be less egregious (and therefore less in need of protection). The Fifth Circuit supplies a concise restatement of the law:

Content-neutral time-place-manner restrictions are examined under intermediate scrutiny, meaning they are permissible so long as they are narrowly tailored to serve a significant governmental interest and leave open ample alternative channels for communication of the information. Content-based time-place-manner restrictions are examined under strict

---

23 This correspondence has not been without criticism. For example, see Simon & Schuster, Inc. v. Members of New York State Crime Victims Bd., 502 U.S. 105, 125 (1991) (Kennedy, J. dissenting).
24 304 U.S. 144, 152 fn. 4 (1938).
scrutiny, meaning they must be narrowly drawn to effectuate a compelling state interest. 26

Things are rather more complex in practice of course. For example, even content-based speech regulation is typically scrutinized under a lower standard when the regulation occurs in a government-owned non-public forum—for example, military bases or prisons. 27 Professor Kelso contends that First Amendment review should therefore be understood as consisting in three tiers: strict scrutiny, intermediate review, and reasonableness balancing. 28 Others have raised doubts about the categorical approach altogether. Bunker, Calvert, and Nevin describe a gradual weakening of strict scrutiny over time, 29 blurring the boundaries of categorical review. Professor Shaman recognizes the trend more broadly in Constitutional jurisprudence, and welcomes a dissolution of “rigid” categories of review. 30

The subtleties of judicial review need not detain us for present purposes. It suffices to observe that even under the courts’ most stringent standard, the courts have held that public policy objectives can justify intrusion on speech rights, so long as those objectives are sufficiently “compelling.” This has consistently remained the majority position among the justices of the U.S. Supreme Court from the earliest First Amendment cases. It should be mentioned however that the defeasibility of speech rights has never enjoyed unanimous approval. Justice Hugo Black envisaged a rather more mechanical approach: that once a category of behavior is recognized as speech, the bar on government intervention should be absolute. 31 Justice Anthony Kennedy, decrying the slipshod manner in which the “compelling state interest” test meandered into First Amendment jurisprudence, expressed similar reservations. 32

Justices William O. Douglas and Antonin Scalia have also been characterized as free speech “absolutists,” although their records are not entirely unambiguous.

In the framework of the harm-based justification, the “absolutist” position appears to be founded upon a rejection of premise (4). I am inclined to agree with this assessment. Even if a speech regulation were shown to be welfare improving, I am skeptical whether it ought to be the government’s business to intervene in the realm of ideas, beliefs, and preferences. Nevertheless, I believe premise (2) to be the more vulnerable step in the harm-based justification. Critical scrutiny should be concentrated upon that inferential move. My purpose in granting premise (4) to proponents of speech regulation is merely to bracket off a digression which might threaten to overwhelm the principal point I hope to advance here: the defectiveness of premise (2).

1.3 Defamation

It is often remarked how few First Amendment cases were litigated prior to the twentieth century. There are several reasons for the relatively delayed emergence of First Amendment case law. Textually, the language of the First Amendment only limits the power of the federal legislature to make laws abridging the freedom of speech.\textsuperscript{33} And the United States Congress evinced little interest in regulating speech in its early years, except in the Alien and Sedition Acts of 1798. The constitutionality of the Alien and Sedition Acts was never litigated before the Supreme Court, as the power of judicial review was not established until 1803,\textsuperscript{34} and the legislation was

\textsuperscript{33}Curiously, “originalist” proponents of speech regulation have relied upon this distinction to justify incursions into the speech realm. Given that the First Amendment is now understood to apply to states under incorporation (see Everson v. Board of Education, 330 U.S. 1 (1947)), the point seems to be that the “intent of the Framers” was only to assign powers in the design of a federal government. Thus, when Madison proclaimed, “The people shall not be deprived or abridged of their right to speak, to write, or to publish their sentiments; and the freedom of the press, as one of the great bulwarks of liberty, shall be inviolable,” the originalists seem to say, he meant only that it should be inviolable \textit{by the federal government}. 1 Annals of Cong. 434 (1789). State governments, of course, may violate at will. And when George Washington, addressing his soldiers, announced, “[T]he freedom of Speech may be taken away, and, dumb and silent we may be led, like sheep, to the Slaughter,” the originalists would contend, he was only talking of the British—not state legislatures. \textit{Founders Online}, National Archives, available at: http://founders.archives.gov/documents/Washington/99-01-02-10840. The contention is baffling.

\textsuperscript{34}Marbury v. Madison, 5 U.S. 137 (1803).
written to expire in 1801.35

The First Amendment would not have protected individuals from state incursions on their speech rights, and the federal government seems not to have tried. Consequently, we have a dearth of First Amendment case law arising from the early decades of the Republic.

However, our historical inquiry need not end here, for the federal constitution is not the sole source of speech rights in American law. Several state constitutions also recognized speech rights, and the justifications for the abridgment of those rights are informative. They are informative insofar as they are examples of rationales which have been offered in defense of speech regulation—not because they have any precedential significance.

The early speech cases typically arose from defamation disputes. It is not difficult to retrieve the harm-deterrence argument implicit in the courts’ opinions. For example, in *Runkle v. Meyer,*36 a printer was sued for publishing a scandalous anecdote involving the putative misdeeds of a clergyman. Among the several issues decided in that case was the question whether libel constituted an actionable claim. Recognizing that “slander” (i.e., a spoken defamation) was actionable, the court reasoned that “libel” (i.e., a printed defamation) ought to be actionable also, on the ground that the harm caused by written defamation, capable of reaching a wider audience, would likely be greater than the harm caused by spoken one.37

The court’s logic relies upon the harm-deterrence argument. Leveraging the harm-of-slander relative to the harm-of-libel in order to establish the actionability of the latter, the court implies that *harmfulness* is the relevant factor in establishing liability. To deploy its argument, the court must assume that slander is prohibited *because* it generates little value and imposes a substantial externality.38 This implies the general principle that the more harmful an expression, the stronger the justification for regulating it. It is only upon this premise that pointing out the relatively

---

35Later Supreme Court jurisprudence suggests that it would—or at least should—have been found unconstitutional. New York Times Co. v. Sullivan, 376 U.S. 254, 576 (1964) (“Although the Sedition Act was never tested in this Court, the attack upon its validity has carried the day in the court of history. Fines levied in its prosecution were repaid by Act of Congress on the ground that it was unconstitutional.”).

363 Yeates 518 (1803).

37Id. at 519 (“The offence of a libel is more heinous, as its circulation of the slander is more extensive . . . .”).

38As to the absence of value in the publication, the court mentions that truth would relieve the printer of his liability. Id. at 520. As to the negative externality, see id. at 519 (“It will not be denied, that if one designedly bespatters another’s cloathes with filth, as he passes the street, . . . he would be liable for damages. And shall a printer with his types, blacken the fairest reputation, the choicest jewel we enjoy . . . .”).
greater harmfulness of printed defamation can have any relevance.

In *Mayrant v. Richardson*, the plaintiff claimed to have lost a congressional election because the defendant had opined to prospective voters, “[Mayrant’s] mind was impaired, weakened, and could never be depended on.” The court ruled that the words did not constitute slander: First, because the extent of the negative externality was thought to be marginal (the court thought the words more likely to inspire “compassion” than “hatred, ridicule, or contempt”). Second, because the expression of an opinion—especially of an individual seeking public office—was at least potentially socially valuable. Thus we find here again, the court’s reasoning relies implicitly upon the harm-deterrence argument: failure to meet conditions (i) and (ii), the utterance was found not harmful and thus not actionable.

*Blunt v. Zuntz* presents a curious context in which an attorney was sued for slander for statements made in court. The court notes:

> It is an established rule, that no action can be maintained against a counsellor for words pertinent to the issue, spoken in judicial proceedings . . . . The object of this rule is closely connected with the *utility of the function of counsel*, which consists principally in a liberal freedom of speech, and that he may not be embarrassed by continually balancing in his mind whether the remark he is about to make be slanderous or not.

The court indicates that the relevant factor was failure of condition (i) to obtain: that the utterances of lawyers in courts are not of low social value. Though it may impose some negative externality, the court recognized the value of such speech to be non-negligible—i.e., there is “utility of the function of counsel.” Ergo, such speech is not harmful, and its expression should not be actionable.

It was not until the mid-twentieth century that the U.S. Supreme Court heard cases challenging defamation claims on First Amendment grounds, yet its justifications were essentially the same as those contemplated in *Runkle, Mayrant, and Blunt.*

---

391 Nott & McC 347 (1818).
40The court stated that for an alleged defamation to be actionable, the externality should either be legal punishment (presumably contemplating false reports to the police or perjury) or measurable injury. *Id.* at 349. The imputation of mental impairment did not, in the view of the court, cause any such injury. *Id.* (“[Mental impairment] was a misfortune and not a fault. It might have been calculated to excite compassion, but not hatred, ridicule or contempt.”); and *id.* at 353 (“[T]he words must be of an opprobrious nature, and such as are calculated to lessen the person of whom they are spoken, in the opinion of the community.”).
41*Id.* at 353.
42Ant.N.P. (2d Ed.) 246.
43*Id.* at fn 1 (emphasis added).
The Court has been explicit in its defamation cases that it is not recognizing a latent definitional quirk of the Constitutional Framers’ conception of “speech,” but rather applying a balancing test.

For example, in New York Times Co. v. Sullivan, the Court ratcheted up protection of speech rights against defamation claims, holding that plaintiffs were required to prove “actual malice” to recover. Critically, like in Blunt, the Court’s reasoning in Sullivan was founded on the failure of condition (i), identifying a non-negligible social benefit in potentially defamatory expressions.

Later, in a distinguishing case limiting the Sullivan holding to public figures, the Supreme Court in Gertz v. Robert Welch, Inc. again identified the balancing of benefits and costs, with Justice Powell writing, “The legitimate state interest underlying the law of libel is the compensation of individuals for the harm inflicted on them by defamatory falsehood.” And in Dun & Bradstreet, Inc. v. Greenmoss Builders, Inc., retreating further from Sullivan, Powell opined, “In light of the reduced constitutional value of speech involving no matters of public concern, we hold that the state interest adequately supports awards of presumed and punitive damages—even absent a showing of ‘actual malice.'”

Indeed, it is abundantly clear from the Court’s language—both when strengthening and weakening the speech right—that it is attempting to identify whether (and to what extent) conditions (i) and (ii) apply. Moreover, it is evident that, to the extent that defamatory speech satisfies the conditions of “harmful speech,” the harm-deterrence argument is assumed. The Court speaks unambiguously in terms of the deterrence objectives—i.e., decreasing the supply of harmful speech—which defamation actions are meant to effect.

---

44 Supra note 35.
45 Sullivan, supra note 35 at 281 (“The importance to the state and to society of such discussions is so vast, and the advantages derived are so great that they more than counterbalance the inconvenience of private persons whose conduct may be involved, and occasional injury to the reputations of individuals must yield to the public welfare, although at times such injury may be great. The public benefit from publicity is so great and the chance of injury to private character so small that such discussion must be privileged.”) (quoting Coleman v. MacLennan, 78 Kan. 711, 724 (1908)).
48 See, e.g., Sullivan, supra note 35, at 279 (“Allowance of the defense of truth, with the burden of proving it on the defendant, does not mean that only false speech will be deterred.”); id. at 299 (“Such criticism cannot [presumably “should not”], in my opinion, be muzzled or deterred by the courts at the instance of public officials under the label of libel.”); Gertz, supra note 46, at 350 (“[Punitive damages for libel] are not compensation for injury. Instead, they are private fines levied by civil juries to punish reprehensible conduct and to deter its future occurrence.”); Dun, supra note 47, at 762 (“[I]n the libel context, the States’ regulatory interest in protecting reputation
1.4 Espionage Act Cases and Incitement

Chronologically, the first significant speech cases litigated under the First Amendment appear in connection with the Espionage Act of 1917 and Sedition Act of 1918. In *Schenck v. United States*, the defendant Charles Schenck, in his capacity as general secretary of the Socialist Party, oversaw the printing and distribution of leaflets encouraging draftees to resist conscription. The government charged that he had violated provisions of the Espionage Act, which prohibited efforts “to obstruct the recruiting and enlistment service of the United States, when the United States was at war with the German Empire.” Writing for a unanimous Court, Justice Holmes devised his infamous “clear and present danger” test:

> The most stringent protection of free speech would not protect a man in falsely shouting fire in a theatre and causing a panic. It does not even protect a man from an injunction against uttering words that may have all the effect of force. The question in every case is whether the words used are used in such circumstances and are of such a nature as to create a clear and present danger that they will bring about the substantive evils that Congress has a right to prevent. It is a question of proximity and degree. When a nation is at war many things that might be said in time of peace are such a hindrance to its effort that their utterance will not be endured so long as men fight and that no Court could regard them as protected by any constitutional right. It seems to be admitted that if an actual obstruction of the recruiting service were proved, liability for words that produced that effect might be enforced.

Holmes’s opinion indicates a balancing of utilities: individuals’ interests in expression and the government’s interest in advancing the war effort. Surmising the utility of the pamphlets negligible relative to the externality, he accordingly finds the legislation constitutional. He is however careful to distinguish that in different circumstances, the calculus of costs and benefits may change. The implication is that the harmfulness of speech—i.e., whether it satisfies conditions (i) and (ii)—is context-dependent, and therefore that the constitutional protection of speech ought to be correspondingly context-dependent.

is served by rules permitting recovery for actual compensatory damages upon a showing of fault. Any further interest in deterring potential defamation through case-by-case judicial imposition of presumed and punitive damages awards on less than a showing of actual malice simply exacts too high a toll on First Amendment values.”) (White, J., concurring in judgment).

49249 U.S. 47 (1919).
50 Id. at 49.
51 Id. at 52.
It is of critical importance to recognize that the conclusion follows only if Holmes assumes the major premise, “If harmful speech is punished, then social welfare is improved.” There is no conceivable other purpose to his pointing out the harmfulness of Schenck’s leafletting except to establish the reduction of harm as a ground for abrogating the speech right. This is further reinforced in Abrams v. United States, the facts of which are substantially similar to those in Schenck, where Holmes wrote in dissent:

I do not doubt for a moment that by the same reasoning that would justify punishing persuasion to murder, the United States constitutionally may punish speech that produces or is intended to produce a clear and imminent danger that it will bring about forthwith certain substantive evils that the United States constitutionally may seek to prevent. The power undoubtedly is greater in time of war than in time of peace because war opens dangers that do not exist at other times.

But as against dangers peculiar to war, as against others, the principle of the right to free speech is always the same. It is only the present danger of immediate evil or an intent to bring it about that warrants Congress in setting a limit to the expression of opinion where private rights are not concerned. Congress certainly cannot forbid all effort to change the mind of the country. Now nobody can suppose that the surreptitious publishing of a silly leaflet by an unknown man, without more, would present any immediate danger that its opinions would hinder the success of the government arms or have any appreciable tendency to do so.

History of course has noted Holmes’s apparently shifting position. In Schenck, he found the wartime leafletting satisfied the requirements of the “clear and present danger” test, while in Abrams, under substantially the same set of facts, he found that it did not. The reason for the apparent transformation has fueled much speculation among historians and biographers. However, Holmes himself contended his position had not changed. The question arises whether Holmes failed to notice the contradictory outcomes arising from the same facts, or whether his protestations of self-consistency were intellectually dishonest. I believe there is a third possibility.

52 250 U.S. 616 (1919).
53 Id. at 627–628 (Holmes, J. dissenting).
54 Holmes, writing for the majority, reaffirmed the “clear and present danger” test in two other pre-Abrams cases, Frohwerk v. United States, 249 U.S. 204 (1919) and Debs v. United States, 249 U.S. 211 (1919). His use of “clear and imminent danger” rather than “clear and present danger” in Abrams is also curious.
55 Id. at 627.
What Holmes may have meant when he claimed his position had not changed was that his reliance on the harm-deterrence argument had remained consistent. In his dissent, Holmes reaffirms the premise that the harmfulness of speech can justify carving out exceptions to the First Amendment. The distinction is merely that in Abrams he no longer seemed to regard anti-war leafletting as satisfying the definition of “harmful speech.” Possibly he thought Schenck’s leaflets more harmful than Abrams’s leaflets, or possibly he reconsidered his assumptions about the harm of leafletting generally. In either case, it would have been the minor premise—not the rule—which determined the different outcomes.

Viewed in the context of Holmes’s jurisprudence generally, it is unsurprising that his First Amendment jurisprudence relied upon distinctly economic premises. The torch was duly passed in United States v. Dennis, in which the great proto-economist judge of the next generation, Learned Hand, elaborated upon the “clear and present danger” test, offering a yet more explicitly economic elaboration: “In each case [courts] must ask whether the gravity of the ‘evil,’ discounted by its improbability, justifies such invasion of free speech as is necessary to avoid the danger.”

Hand’s analysis is uncannily familiar, for it closely resembles his negligence formula in United States v. Carroll Towing Co., ubiquitously cited in torts casebooks, which prescribes a balancing of expected benefits against expected costs to effect efficient incentives. The Dennis elaboration of the “clear and present danger” test is essentially an extension of the tort principle to the First Amendment context.

Interestingly, prior to Dennis, when Hand was still a district court judge, he had fashioned a rather different First Amendment test in Masses Publishing Co. v. Patten. The plaintiff—publisher of the socialist magazine The Masses—moved for a preliminary injunction against the postmaster, who refused to deliver the publication on the ground that it was “seditious” under the terms of the Espionage Act. The magazine was critical of the decision to go to war and praised citizens who would resist conscription. Hand’s test in Masses was to require that the incitement be direct in order to justify exception to the general speech right.

---

56 United States v. Dennis, 183 F.2d 201, 212 (2d Cir. 1950), aff’d, 341 U.S. 494 (1951).
57 159 F.2d 169 (2d. Cir. 1947) (defining the duty of care as $B < PL$, or the burden of precautions less than the probability of loss). Posner, of course remarks upon the convergence of principle. Free Speech, supra note 7 at 8; Economic Analysis of Law, supra note 7 at 958.
58 244 F. 535 (1917).
59 Id. at 540 (“[T]o assimilate agitation, legitimate as such, with direct incitement to violent resistance, is to disregard the tolerance of all methods of political agitation which in normal times is a safeguard of free government. The distinction is not a scholastic subterfuge, but a hard-bought acquisition in the fight for freedom, and the purpose to disregard it must be evident when the power exists. If one stops short of urging upon others that it is their duty or their interest to resist the
prohibiting the expression, “You must violently overthrow the government,” would be constitutional, whereas prohibition of the expression, “It would be laudable if someone would overthrow the government,” would not be. The distinction is a crystalline one, for it reduces the factual question to one of linguistics. The former expression is imperative, whereas the latter is declarative.

Throughout his *Masses* opinion, Hand repeatedly emphasized the high social value of political dissent as a ground for narrowly construing the Espionage Act. Thus, even before the question went to the Supreme Court in *Schenck*, it seems Hand had already envisaged that conditions (i) and (ii) would need to be satisfied to overcome the general speech right. We thus see a kind of convergent basis for rule-making in the speech context. Ultimately, Hand granted the plaintiff’s motion for preliminary injunction on the ground that condition (i) was not satisfied. The Second Circuit reversed, and the Supreme Court, as we have seen, went a different way—at least initially. Hand’s test in *Masses* is thought to have inspired the “imminent lawless action” test in *Brandenburg v. Ohio*, which later displaced the “clear and present danger” test.

Regardless, there can be little doubt that the courts were, even in their nascent contemplation of the First Amendment, already assuming the harm-deterrence argument in their reasoning.

### 1.5 Fighting Words and Hate Speech

A third context in which the courts have recognized limitations to the general speech right is the “fighting words” exception. The precedent was established in *Chaplin-sky v. New Hampshire*, in which Justice Murphy, writing for a unanimous Court, premised the constitutionality of speech regulations upon whether the targeted expressions, “by their very utterance inflict[ed] injury or tend[ed] to incite an immediate breach of the peace.”

In that case, a proselytizing Jehovah’s Witness was fined for law, it seems to me one should not be held to have attempted to cause its violation.”).

---


61 395 US 444 (1969). Although *Brandenburg* remains the standard for judging the constitutionality of incitement laws today, the published opinions are not particularly enlightening for present purposes. Chief Justice Warren’s majority opinion gives no reason for the departure from the Holmes cases, pretending continuity with *Dennis*. Justice Black’s concurring opinion, consisting in one paragraph, merely indicates his view that *Brandenburg* does invalidate the “clear and present danger” test. And Justice Douglas’s concurrence takes the position that the First Amendment should not be subject to any exceptions, and thus supplies no reasoning to justify why the Court carves out the limits that it does.

insulting a police officer, calling him a “damned racketeer,” and a “damned Fascist,” in violation of New Hampshire laws prohibiting expressions susceptible to provoking violence. In justifying the holding, Murphy explained:

It has been well observed that such utterances are no essential part of any exposition of ideas, and are of such slight social value as a step to truth that any benefit that may be derived from them is clearly outweighed by the social interest in order and morality. 63

The identification of conditions (i) and (ii) could not more clearly have been within the contemplation of the Court. Moreover, the harm-deterrence argument is also evident in the Court’s reasoning. Envisioning a deterrent effect leading to an improvement in social welfare, Murphy wrote:

[T]he statute had been previously construed as intended to preserve the public peace by punishing conduct, the direct tendency of which was to provoke the person against whom it was directed to acts of violence. 64

The hate speech cases, drawing upon Chaplinsky, are an interesting source of further discussion. In R.A.V. v. City of St. Paul, 65 a curiously divided Court—albeit unanimous in judgment—ruled upon the constitutionality of a municipal ordinance, proscribing expressive activities which have the propensity to arouse “anger, alarm or resentment in others on the basis of race, color, creed, religion or gender.” 66

The petitioner was charged under the St. Paul ordinance for burning a cross in the front yard of a Black family. Justice Scalia, writing for the majority, premised his judgment upon an impermissible motive of the legislators. The gravamen of his argument was that even where legislation falls within a recognized exception to the general speech right, prohibiting only a subset of that excepted class requires independent justification. Scalia’s concern was evidently that Chaplinsky could be used as a workaround to allow content-based regulations.

That would mean that a city council could enact an ordinance prohibiting only those legally obscene works that contain criticism of the city government or, indeed, that do not include endorsement of the city government. . . . The proposition that a particular instance of speech can be proscribable on the basis of one feature (e.g., obscenity) but not on

---

63 Id. (emphasis added).
64 Id. at 574, fn 8.
66 Id. at 377.
the basis of another (e.g., opposition to the city government) is commonplace and has found application in many contexts. We have long held, for example, that nonverbal expressive activity can be banned because of the action it entails, but not because of the ideas it expresses—so that burning a flag in violation of an ordinance against outdoor fires could be punishable, whereas burning a flag in violation of an ordinance against dishonoring the flag is not.\textsuperscript{67}

Scalia’s apprehension seems to be a legitimate one, which the concurring opinions may not have fully appreciated. Nevertheless, Scalia rested his argument upon this indeterminate observation alone. Evidently, he took \textit{Chaplinsky} to do no more than carve out a constitutional exception for the category “fighting words.” Yet \textit{Chaplinsky} does more than this. It establishes a cost-benefit test. Having determined—sensibly—that the abrogation of speech rights within the subset of a recognized exception requires independent justification, Scalia does not proceed to the inquiry whether the expressions proscribed under the St. Paul ordinance satisfy the \textit{Chaplinsky} test. Rather, he strikes it down on the ground that it is content-based.

The concurring judgments of Justices White, Blackmun, and Stevens are unspARING in their criticism of the majority. White, curiously, objects to Scalia for requiring an independent justification, contending that speech within a recognized exception may permissibly be proscribed without further inquiry.\textsuperscript{68} Blackmun’s concurrence is ambiguous, on the one hand reasoning that if the regulation of fighting words is permissible, then its more harmful variants ought—on the basis of being more harmful—be regulable \textit{a fortiori};\textsuperscript{69} and on the other hand, that the parsing of recognized exceptions will lead to a dissolution of speech protections.

Stevens’s concurrence seems to provide the widest perspective, evincing an appreciation of the principles motivating both Scalia and White,\textsuperscript{70} and locating a sensible role for the \textit{Chaplinsky} test and categorical reasoning. Specifically, Stevens favored using harm-based justifications to support categorical exclusions to the general speech right. The mechanics of how such a review would operate are not however entirely clear from his opinion.

\begin{itemize}
\item\textsuperscript{67} Id. at 384.
\item\textsuperscript{68} Id. at 400 ("To the contrary, those statements meant precisely what they said: The categorical approach is a firmly entrenched part of our First Amendment jurisprudence. . . . [T]he majority holds that the First Amendment protects those narrow categories of expression long held to be undeserving of First Amendment protection—at least to the extent that lawmakers may not regulate some fighting words more strictly than others because of their content.").
\item\textsuperscript{69} Echoing the argumentative move in \textit{Runkle, supra} note 36.
\item\textsuperscript{70} 505 U.S. at 417.
\end{itemize}
R.A.V. is a chaotic case. The justices—Stevens excepted—seem to some extent to have been talking past one another. Most relevant to the present inquiry, Scalia’s opinion eschews harm-based arguments entirely. His opinion proposes a purely legalistic test: \( x \) is a permissible regulation of speech if and only if \( x \) is neither more nor less than the whole category of obscenity, fighting words, criminal conspiracy, defamation, etc. The reasons why these categories have been carved out of the general speech right are given no serious consideration. Scalia declines to articulate whether additional categories should be added, or whether changing circumstances might warrant removing a category.\(^{71}\) The basis is characterized in formalistic terms (patterned on “legal logic” rather than consideration of consequential effects). The concurring opinions make much of this omission, with White writing:

This selective regulation reflects the city’s judgment that harms based on race, color, creed, religion, or gender are more pressing public concerns than the harms caused by other fighting words. In light of our Nation’s long and painful experience with discrimination, this determination is plainly reasonable. Indeed, as the majority concedes, the interest is compelling.\(^{72}\)

Blackmun’s concurrence also mentions the relative harmfulness of generic fighting words as compared with racially charged symbols.\(^ {73}\) And Stevens discusses at length the harmfulness justification in limiting First Amendment rights in several contexts.\(^ {74}\)

Scalia’s opinion is interesting because it represents one of the few genuine exceptions where the Court has employed non-harm-based justifications for speech regulation. It is important to recognize however that in R.A.V., the result was not to allow the disputed regulation, but rather to invalidate it. Discerning a consistent thread in Scalia’s First Amendment jurisprudence is of course a difficulty which has motivated much commentary, and it is unclear to what extent Scalia’s opinion in R.A.V. should be taken to conflict with the economic arguments undergirding Chaplinsky.\(^ {75}\) Regardless, taken in the context of succeeding cases, R.A.V. seems not to

---

\(^{71}\)Scalia acknowledges that “compelling state interests” may override the general speech right, conceding to cost-benefit weighing in principle, though he seems to regard the St. Paul ordinance as overly broad in achieving the government’s objective. *Id.* at 395–96.

\(^{72}\)Id. at 407.

\(^{73}\)Id. at 415.

\(^{74}\)Id. at 423, 424–26, 429–435.

\(^{75}\)Justice Scalia’s First Amendment jurisprudence is a fascinating web of apparently contradictory positions, which I cannot satisfactorily explore within the scope of this article. On the one hand, he established a reputation as a stalwart defender of individual speech rights in cases like R.A.V., *id.*; Texas v. Johnson, 491 U.S. 397 (1989) (finding prohibitions on flag-burning unconstitutional); and
represent a general departure from harm-based justifications for speech regulation for the Court as a whole.

In a factually similar subsequent case, *Virginia v. Black,* the Court struggled in its application of Scalia's content/category test from *R.A.V.* Justice O'Connor, writing for the majority, distinguished that the Virginia statute at issue, which prohibited cross-burning specifically, was constitutional under the “true threats” exception—though the specification of cross-burning as prima facie evidence of intimidation was not constitutional (being “content-based”). As Justice Souter observes in his dissent, distinguishing the prima facie designation from the substantive prohibition cannot sensibly save the statute, for the prohibition itself (not merely the presumption clause) specifies *cross-burning* rather than “true threats” generally. O'Connor’s response was to distinguish Virginia’s statute—singing out cross-burning—as consistent with *R.A.V.*, because cross-burning is an especially virulent form of intimidation. It constituted a permissible partition of a recognized exception, because


77 VA Code Ann. § 18.2–423 states:

> It shall be unlawful for any person or persons, with the intent of intimidating any person or group of persons, to burn, or cause to be burned, a cross on the property of another, a highway or other public place. Any person who shall violate any provision of this section shall be guilty of a Class 6 felony.

> Any such burning of a cross shall be prima facie evidence of an intent to intimidate a person or group of persons.

78 *Id.* at 363 (“It shall be unlawful for any person or persons, with the intent of intimidating any person or group of persons, to burn, or cause to be burned, a cross on the property of another, a highway or other public place. Any person who shall violate any provision of this section shall be guilty of a Class 6 felony. Any such burning of a cross shall be prima facie evidence of an intent to
the *reason* for regulating the subset was identical to the *reason* for regulating the category generally. The distinction seems a flaccid one, for one might just as well say that the St. Paul ordinance was directed at a subset of “fighting words,” which are especially virulent in the form of a burning cross.

Regardless how the cross-burning cases should be reconciled, or how the Court would decide future cross-burning cases, the relevant point is that to the extent that *R.A.V.* seemed to eschew the harm-based justification, *Black* grafted it onto the content/category test in allowing “particularly virulent” subsets to be regulable independently of a category. And even *R.A.V.* were understood to represent an abandonment of the harm-based justification, it was reinstated in *Black*.

One last representative case, which may be worth considering in the context of fighting words and hate speech, is *Snyder v. Phelps*, where members of the infamous Westboro Baptist Church picketed at the funeral of a fallen marine with signs reading, “God hates fags,” and “Fags doom nations.” They were sued by the soldier’s family for intentional infliction of emotional distress. Writing for the majority, Justice Roberts recognized the distasteful speech to be constitutionally protected, emphasizing the non-negligible value of political speech:

> The “content” of Westboro’s signs plainly relates to public, rather than private, matters. The placards highlighted issues of public import—the political and moral conduct of the United States and its citizens, the fate of the Nation, homosexuality in the military, and scandals involving the Catholic clergy—and Westboro conveyed its views on those issues in a manner designed to reach as broad a public audience as possible. Even if a few of the signs were viewed as containing messages related to a particular individual, that would not change the fact that the dominant theme of Westboro’s demonstration spoke to broader public issues.

Restated in the framework of harm-based justification, Roberts’s opinion indicates the failure of condition (i). Because condition (i) fails to obtain, the speech is not “harmful speech,” therefore the harm-based justification fails, and therefore there is no exception to the general speech right protecting it.

Justice Alito, the lone dissenter, engages the majority on precisely these grounds, citing *Chaplinsky* and asserting the satisfaction of condition (i): “Allowing family

---


80 *Id.* at 469. It is perhaps worth mentioning, given the theme of the signage, that the deceased was not gay. Evidently, members of the Westboro Baptist Church inferred his death to have been vengeance for what they perceived to be a generalized moral deterioration of Christian values.

81 *Id.* at 444.
members to have a few hours of peace without harassment does not undermine public debate”;82 and also satisfaction of condition (ii): that the picketing would have the “potential to wound as a personal verbal assault on a vulnerable private figure.”83

Concededly, I have given short shrift to of the doctrinal distinctions which the Court (and Alito’s dissent) purports to have motivated its decision. That is, the distinction between speech relating to matters of public versus private concern. The doctrine is, I think, little more than a categorical presumption relating to the conditions of harmful speech. Pulling on the thread, it reveals itself to align neatly with the harm-based justification, relying on the additional premise that matters of public concern presumptively fail condition (i). The argument for the presumption is that matters relating to politics and social institutions are so paramount in importance, that any abrogation of that type of speech should be especially disfavored.84 Put differently, the public/private distinction is premised upon the premise that political speech (broadly construed) is so important that it should be singled out as failing condition (i) by default.85

82 Id. at 473.
83 Id. at 475.
84 The Court’s unwavering commitment to political speech over other categories seems myopic. I would hazard to suppose that if violinists were polled what kinds of expression deserved the most stringent protection, they might name music; and if painters, then paintings; and if physicists, then physics publications. The privileging of political speech seems to be founded upon little more than the vanity of political actors. Looking across the broad sweep of history, I can think of no statesman, activist, revolutionary, or judge whose expressive contributions compare favorably to those of Bach, Shakespeare, Picasso, or Newton. And I see no sensible reason why political speech should be privileged over expressions in art, science, mathematics, or philosophy. Presumably, the thought is that political speech is most important, because it is essential to a well-functioning liberal democracy. And poorly functioning social institutions—especially those of an autocratic or oppressive government—might impinge upon all other areas of speech. Thus, if any category of speech is valuable, then political speech must be the most valuable of all. See cases cited supra note 85. This argument is weak. Even if the premise were conceded, the form of the argument is to privilege means over ends. It is like claiming that instruments are more valuable than music, because without instruments there would be no music. How unquestioningly courts seem to accept this plainly fallacious reasoning is nothing short of baffling. Yet regardless whether this is right, the point is that the Court has viewed political speech as deserving special protection because it regards expressions of political speech as being distinctly unlikely to be expressions which satisfy condition (i).
85 See, e.g., Monitor Patriot Co. v. Roy, 401 U.S. 265, 272 (1971) (“[It] can hardly be doubted that the constitutional guarantee has its fullest and most urgent application precisely to the conduct of campaigns for political office.”); Whitney v. California, 274 U.S. 357, 375 (1927) (Brandeis, J. concurring) (“Those who won our independence believed that the final end of the state was to make men free to develop their faculties, and that in its government the deliberative forces should prevail over the arbitrary. . . . They believed that freedom to think as you will and to speak
1.6 Commercial Speech

Another realm of expression where the Court has commonly allowed regulation is “commercial speech.” The rationale tracks the “public versus private concern” distinction previously discussed, which in turn hinges upon the satisfaction of condition (i).

As a historical matter, the Courts’ development of its commercial speech doctrine was erratic. In Valentine v. Chrestensen, the Court held, “We are . . . clear that the Constitution imposes no . . . restraint on government as respects purely commercial advertising.”\(^{86}\) The Court cites no precedents, nor provides any argument for the principle. Subsequent scholarship and judicial opinions have regarded the decision unfavorably.\(^{87}\)

In truth, the doctrine was hardly “clear” in the common law of the states prior to Valentine. For example, in People v. Osborne,\(^{88}\) a California court ruled that the state could not permissibly interfere in the speech of a barber, who was fined for displaying prices for his services outside his shop in violation of a city ordinance.\(^{89}\) The court observed:

The constitutional liberty of speech implies the right to freely utter and publish whatever the citizen may please and immunity from legal censure and punishment for the publication so long as it is not harmful in its

as you think are means indispensable to the discovery and spread of political truth; that without free speech and assembly discussion would be futile; that with them, discussion affords ordinarily adequate protection against the dissemination of noxious doctrine; that the greatest menace to freedom is an inert people; that public discussion is a political duty; and that this should be a fundamental principle of the American government.”); Mills v. State of Alabama, 384 U.S. 214, 218–219 (“Whatever differences may exist about interpretations of the First Amendment, there is practically universal agreement that a major purpose of that Amendment was to protect the free discussion of governmental affairs. This of course includes discussions of candidates, structures and forms of government, the manner in which government is operated or should be operated, and all such matters relating to political processes.”).

\(^{86}\)316 U.S. 52, 54 (1942).


\(^{89}\)Predating incorporation of the First Amendment (See Carolene Products, 304 U.S. 144 (1938) and Everson, 330 U.S. 1 (1947)), Osborne was decided in the context of the speech guarantee of the California Constitution.
character when tested by such standards as the law affords, and what may be spoken may be written.\(^90\)

Similarly, commercial speech in the context of newspaper advertising was recognized in earlier cases as protected speech.\(^91\) Yet regardless, \textit{Valentine} was the first direct treatment of “commercial speech” by the U.S. Supreme Court in relation to the federal constitution. And the Court proclaimed no protection whatever to its expression.

Doubts about the doctrine grew steadily in the ensuing decades.\(^92\) \textit{Valentine} was finally overturned in \textit{Virginia State Pharmacy Board v. Virginia Citizens Consumer Council},\(^93\) which concerned a statute declaring it unprofessional for pharmacists to advertise prices, fees, premiums, discounts, rebates, or credit terms.\(^94\) Justice Blackmun, writing for the majority, rejected the proposition that commercial speech presumptively satisfied the conditions of harmful speech:

> Our question is whether speech which does no more than propose a commercial transaction is so removed from any exposition of ideas, and from truth, science, morality, and arts in general, in its diffusion of liberal sentiments on the administration of Government, that it lacks all protection.
> Our answer is that it is not.\(^95\)

It is revealing to observe the \textit{way} that Blackmun attacks the commercial speech exception established in \textit{Valentine}. The form of his argument is straightforward. He disputes whether commercial speech is presumptively harmful speech, explaining how plausible it would be for instances of commercial speech to fail conditions (i) and (ii). Describing why it ought not be assumed that commercial speech satisfies condition (i):

> So long as we preserve a predominantly free enterprise economy, the allocation of our resources in large measure will be made through numerous

\(^{90}\)59 P.2d at 1087 (\textit{citations omitted}).

\(^{91}\)See, \textit{e.g.}, Ex parte Jackson, 96 U.S. 727, 733 (1878); In re Rapier, 143 U.S. 110, 134 (1892); Lewis Publishing Co. v. Morgan, 229 U.S. 288, 315 (1913).


\(^{93}\)425 U.S. 748 (1976).


\(^{95}\)425 U.S. at 762 (\textit{citations omitted}).
private economic decisions. It is a matter of public interest that those decisions, in the aggregate, be intelligent and well informed. To this end, the free flow of commercial information is indispensable. . . . even if the First Amendment were thought to be primarily an instrument to enlighten public decisionmaking in a democracy, we could not say that the free flow of information does not serve that goal.96

And why it ought not be assumed that commercial speech satisfies condition (ii):

Price advertising, it is said, will reduce the pharmacist’s status to that of a mere retailer. The strength of these proffered justifications is greatly undermined by the fact that high professional standards, to a substantial extent, are guaranteed by the close regulation to which pharmacists in Virginia are subject. And this case concerns the retail sale by the pharmacist more than it does his professional standards. Surely, any pharmacist guilty of professional dereliction that actually endangers his customer will promptly lose his license.

Subsequent courts have repeated in unequivocal terms the proposition that the justification for regulating commercial speech is harm-based. For example, in City of Cincinnati v. Discovery Network, Inc.,97 the Supreme Court wrote, “[The government] has not asserted an interest in preventing commercial harms by regulating the information distributed by respondent publishers’ newsracks, which is, of course, the typical reason why commercial speech can be subject to greater governmental regulation than noncommercial speech.”98

Analysis of an intermediate standard of review was articulated four years later in Central Hudson Gas & Electric Corp. v. Public Service Commission,99 where the court announced a four-part test for determining the constitutionality of commercial speech regulations. Only the first element is relevant to the present inquiry—i.e., that the regulation of commercial speech which is “misleading” merits no heightened scrutiny.100

The proposition that misleading statements or “false advertising” should be regulable is grounded upon its harmfulness.101 The assumption is that false or misleading utterances have a systematic tendency to generate negligible social value—there

---

96 Id. at 765.
98 Id. at 426.
100 Id. at 564.
101 See, e.g., Bolger v. Youngs Drug Products Corp., 463 U.S. 60, 81 (1983) (“The interest in protecting consumers from commercial harm justifies a requirement that advertising be truthful
being no good in promoting false beliefs—and a substantial negative externality arising from the disappointment or injury caused to consumers. Joined with the harm-deterrence argument, the received view asserts, when a category of speech satisfies conditions (i) and (ii), its regulation is justified. Therefore, regulating commercial speech with the objective of deterring false advertising enjoys the support of a harm-based justification.

1.7 Obscenity

The last speech exception, for which I furnish examples of harm-based justification, is obscenity. Here, the appeal to the harmfulness of the proscribed expression is relatively undisguised.

Obscene speech was of course commonly regulated in English law. Whether this practice was intended to carry over after ratification of the Bill of Rights was an open question. However, as I discuss above, the United States Congress was in the early years of the republic generally disinclined to intrude upon speech—with the exception of the short-lived Alien and Sedition Acts—and there were no federal obscenity laws until the mid-nineteenth century. Wholesale regulation of obscenity was not aggressively pursued until passage of the Comstock Act in 1873, which banned the mailing of “[e]very obscene, lewd, or lascivious, and every filthy book, pamphlet, picture, paper, letter, writing, print, or other publication of an indecent . . . .”}; Procter & Gamble Co. v. Amway Corp., 242 F.3d 539, 549 (5th Cir. 2001) (“[F]alse or misleading commercial speech should receive no protection, because commercial speech merely gives information to consumers about a producer’s goods, and any false information either has no value or is harmful.”); Bates v. State Bar of Arizona, 433 U.S. 350, 379 (1977) (justifying regulation of false advertising by attorneys).

102 See, e.g., Bates, 433 U.S. at 380 (acknowledging the balancing of “possible harm to society from allowing unprotected speech to go unpunished”); Edenfield v. Fane, 507 U.S. 761, 762 (1993) (“A governmental body seeking to sustain a restriction on commercial speech must demonstrate that the harms it recites are real and that its restriction will in fact alleviate them to a material degree.”).

103 Central Hudson, 477 U.S. at 563 (“The First Amendment’s concern for commercial speech is based on the informational function of advertising. Consequently, there can be no constitutional objection to the suppression of commercial messages that do not accurately inform the public about lawful activity. The government may ban forms of communication more likely to deceive the public than to inform it.”) (citations omitted).


105 CITE
character,” as well as any information or products which may be used for the purpose of contraception.

To be sure, the legislation’s leading proponent, Anthony Comstock, offered only vague consequentialist justifications for regulating speech. Victorinan social norms, a prying preoccupation with the sexual activities of the others, and assorted moral and religious fixations seem to have constituted Comstock’s primary motivations. His moral crusading might be regarded quaint today, if not for a significant contingent of religious ideologues committed to carrying his torch into the twenty-first century. Unfortunately, Comstock’s prurient grandstanding should not be dismissed as merely comic, for moral crusaders of his ilk still exert some—if thankfully dwindling—influence in the political arena.

Like his modern-day successors, the nearest Comstock came to a consequentialist justification for the wide-ranging regulation of “morally corrupting” speech was to gesture vaguely at the corruption of children and dissolution of the family. It is difficult to address these amorphous speculations directly, for neither Comstock nor his successors have attempted anything resembling a rigorous argument for them. He merely insists—without evidence—upon dubious psychological conjectures.


See Anthony Comstock, Traps for the Young (4th ed., 1883). This work is a singularly fascinating historical artifact, which I would commend anyone interested in obscenity law to skim. It is a tour de force of moral righteousness and paranoia: Comstock wastes no time, writing from the very first chapter, “Evil thoughts, like bees, go in swarms. . . . There is something wonderfully strange in the rapidity with which youthful minds take up lewd thoughts and suggestions.” Id. at 7.

In a later chapter discussing the “free-love trap” (essentially, all sexual relationships except marriage), Comstock analogizes pre-marital and extra-marital sex to a “stone trap” he devised as a child to crush squirrels and rabbits. The analogy may perhaps resonate with priggish psychopaths—I must confess the significance of the grotesque allusion escapes me. Of “free love,” he preaches:

It takes the word ‘love,’ that sweetens so much of earth, and shines so brightly in heaven, and making that its watchword, distorts and prostitutes its meaning, until it is the mantle for all kinds of license and uncleanness. It should be spelled l-u-s-t, to be rightly understood, as it is interpreted by so-called liberals. . . . As advocated by a few indecent creatures calling themselves reformers—men and women foul of speech, shameless in their lives, and corrupting in their influences—we must go to a sewer that has been closed, where the accumulations of filth have for years collected, to find a striking resemblance to its true character. I know of nothing more offensive to decency, or more revolting to good morals, than the class of publications issuing from this source.

Id. at 158. Though Anthony Comstock’s personage has faded somewhat from public notoriety, I
I shall nevertheless attempt a rational reconstruction of Comstock’s argument: it seems he takes as a premise that human beings are universally happier when partaking in lifelong monogamous relationships, spawning offspring, and practicing obeisance to religious norms. He also evidently believed that ordinary persons are liable to be tempted—easily tempted—to a less fulfilling existence mired in sin and debauchery. The law ought therefore intervene, to set children and wayward adults on a better path, obliterating any information which might entice the weak-willed to act against their self-interest.

Though I have attempted to render Comstock’s claims as charitably as I think possible, it is difficult to take any step of the argument seriously. It is plainly a harm-based justification, though for reasons I will enumerate below, I am skeptical whether it was (or is) ever intended sincerely. Although my aim in this article is to rebut harm-based justifications for the regulation of speech, and Comstock’s argument is a harm-based one, I should not want to lump Comstock’s argument with the rest. It is especially bad, and warrants separate treatment.

First, there is no evidence that human beings are generally (much less universally) happier in lifelong monogamous relationships. The archaeological record strongly suggests that early humans were polyamorous. Lifelong monogamous relations are an artifact of culture—not an inherent biological tendency. The preference for monogamy—to the extent that it exists—is not innate in the human species.

Of course, biology alone is not dispositive. Family values advocates will eagerly point to studies measuring the relative “happiness” of married versus unmarried persons, which they claim as evidence that traditional “family oriented” lifestyles tend to be more fulfilling. However, these surveys fail to disambiguate the influence of social norms and cultural expectations. Activists misread the research as establishing the proposition that people tend to be happier in marriages. However, the data could just as well be read to reveal the perniciousness of social norms and expectations: that nonconformity with conventional social practices results in decreased happiness. In other words, the data could be read as indicating not that marriage makes people happier, but rather that intolerant pro-marriage cultures cause unmarried people feel less happy. Research on the discriminatory treatment of non-married people is still at a nascent stage, yet the evidence of insidious and pervasive

hazard to suppose historians shall one day yet accord him his rightful place alongside the great moral purifiers—Savonarola, Torquemada, and Robespierre.

108CITE MORE  See generally Christopher Ryan & Cacilda Jethá, Sex at Dawn: How We Mate, Why We Stray, and What It Means for Modern Relationships (Harper 2011).

109CITE

110Moreover, no studies of which I am aware control for omitted variable bias.
mistreatment is compelling. The case for procreation as a universal private good is weaker still. Even in the face of pervasive social pressure to bear and raise children, there exists strong empirical evidence that individuals with children tend to be less happy than those without. More probing research on the effects of children on the welfare of parents is suspiciously difficult to find. Yet given the opportunity costs associated with having children, it is plausible to suppose that parents also suffer—in addition to psychological detriment—significant financial disadvantages from having children. Comstock’s premise—that people tend to be “better off” having children—is wholly contradicted by the evidence. There may of course be other public policy arguments for encouraging procreation. I am doubtful however that anyone could reasonably argue that procreation is better for parents.

Finally, even if we were to grant arguendo that people tended to be more satisfied in lifelong monogamous relationships producing children, there is no evidence for the causal inference that individuals are as easily led to deviate from welfare-maximizing choices as Comstock assumes. The Comstock argument presupposes of humankind a degree of such abject stupidity, that even the hint of an “unwholesome” thought would entice countless innocents to self-destruction. Even conceding that human decision-making is frequently and systematically susceptible to cognitive bias, there is no principled reason to suppose “unwholesome” choices to be the product of cognitive bias rather than straightforward preference satisfaction. If lifelong monogamous relationships producing children were so clearly utility-maximizing, then the law would not need to enforce it. People would gladly pursue it of their own accord.

Curiously, Comstock denied that the regulation of obscene speech was speech regulation at all. In a glib move—frequently rehearsed by his successors—Comstock claimed that under his namesake legislation citizens would be free to express any thought or idea they pleased, so long as the expression were a legally permissible one. Thus, it was no abrogation of the speech right to censor obscene expressions. It amounts to little more than saying that people are free to express what they please, except when they are not. Obviously, by this reasoning, the freedom of expression would be operative under any set of laws, no matter how restrictive. It is astounding


112 CITE.

how frequently this patently circular rhetoric is repeated.

Regardless of this, I expect that the modern day proponents of Comstock’s position will find my counterarguments unsatisfying, because the harm-based arguments they propose were never what truly motivated their position. The consequentialist argument is for them an afterthought. The true ground for Comstockery seems not to be the regulation of harm, but rather the enforcement of morals. The reason why I expect my counterarguments against their harm-based justification are unlikely to persuade them is because the harm-based justification was never the real justification for their position.

Nevertheless, it is revealing that Comstock and his successors felt the need to construct a consequentialist facade. They seem to intuit that nakedly advocating for the enforcement of morals would fail to persuade legislators, judges, and the public. There is implicit in their pseudo-harm-based justification a recognition that the law would require an identification of harms.

The earliest court opinions ignored the First Amendment implications of the Comstock Act.\textsuperscript{114} Disputes about its enforcement hinged upon procedural issues and precisification of what kinds of materials constituted obscene expressions.\textsuperscript{115} The test which emerged was imported from the British case, \textit{Regina v. Hicklin}.\textsuperscript{116} For an expression to be obscene, it must “deprave and corrupt those whose minds are open to such immoral influences, and into whose hands a publication of this sort may fall.”\textsuperscript{117} Deferring to the legislature, the courts refrained from undertaking any constitutional review of the justification for the obscenity exception. To the extent the courts averred any justification at all, it was to acknowledge the propensity of obscene material to “corrupt the morals” of the citizenry (it is baffling that the courts

\textsuperscript{114}See, \textit{e.g.}, United States v. Bennett, 24 F.Cas. 1093 (1879); Rosen v. United States, 161 U.S. 29 (1896).

\textsuperscript{115}Skepticism about the semantic project was immediate and widespread, with Judge J.C. Ruppenthal neatly summing up:

\begin{quote}
From the foregoing it may be seen that no general principle runs through the statutes of all the states, etc. As with laws everywhere that impinge upon sex matters in any way, there is more of tabu and superstition in the choice and chance, the selection and caprice, the inclusions and exclusions of these several enactments than any clear, broad, well-defined principle or purpose underlying them. Without such principle, well-defined and generally accepted, the various laws must remain largely haphazard and capricious.
\end{quote}


\textsuperscript{116}L. R. 3 Q. B. 360 (1868).

\textsuperscript{117}Bennett, 24 F.Cas. at 1104 (quoting Hicklin).
seem unquestioningly to have regarded this the business of government to do).\footnote{Judge Blatchford uses the phrase “corruption of morals” fourteen times in Bennett. \textit{Id.}}

Such a condition could not stand, of course, and the courts have subsequently devised harm-based justifications—albeit ex post facto—to support the obscenity exception. The first test was articulated in \textit{Roth v. United States}.\footnote{354 U.S. 476 (1957).} Writing for the majority, Justice Brennan invalidated the \textit{Hicklin} test writing:

\begin{quote}
The Hicklin test, judging obscenity by the effect of isolated passages upon the most susceptible persons, might well encompass material legitimately treating with sex, and so it must be rejected as unconstitutionally restrictive of the freedoms of speech and press.\footnote{\textit{Id.} at 489.}
\end{quote}

Brennan articulated the new standard: “whether to the average person, applying contemporary community standards, the dominant theme of the material taken as a whole appeals to prurient interest.”\footnote{\textit{Id.}}

In justifying the new standard, Brennan seems to reject the need for harm-based justification, writing:

\begin{quote}
It is insisted that the constitutional guaranties are violated because convictions may be had without proof either that obscene material will perceptibly create a clear and present danger of antisocial conduct, or will probably induce its recipients to such conduct. But, in light of our holding that obscenity is not protected speech, the complete answer to this argument is in the holding of this Court in \textit{Beauharnais v. People of State of Illinois}, [343 U.S. 250, 266 (1952).]

Libelous utterances not being within the area of constitutionally protected speech, it is unnecessary, either for us or for the State courts, to consider the issues behind the phrase “clear and present danger.” Certainly no one would contend that obscene speech, for example, may be punished only upon a showing of such circumstances.\footnote{\textit{Id.} at 486–487.}
\end{quote}

Yet while denying the requirement of a harm-based justification, he contradictorily proceeds to reason in terms of harm-based justification. Brennan first asserts that exceptions to the general speech right must satisfy condition (i), writing:

\begin{quote}
\footnote{36}
All ideas having even the slightest redeeming social importance—unorthodox ideas, controversial ideas, even ideas hateful to the prevailing climate of opinion—have the full protection of the guaranties, unless excludable because they encroach upon the limited area of more important interests. But implicit in the history of the First Amendment is the rejection of obscenity as utterly without redeeming social importance. This rejection for that reason is mirrored in the universal judgment that obscenity should be restrained . . . .

And with respect to the particular controversy of the case, he emphasizes the likely failure of condition (i) to obtain, writing:

Sex, a great and mysterious motive force in human life, has indisputably been a subject of absorbing interest to mankind through the ages; it is one of the vital problems of human interest and public concern.

Thus, despite asserting the absence of a harm-based justification, Brennan justifies invalidating the Hicklin test on the basis of overbreadth for likely failure of conditions (i) and (ii) to obtain in a substantial number of instances.

The cases which followed reveal the courts struggling to make sense of the Roth standard. The confusion was largely resolved in Miller v. California, adopting an unequivocally harm-based justification for the obscenity exception, and setting the standard which remains the basis for determining whether expressions fall under the obscenity exception. Justice Burger, writing for the majority, established the three-part test:

The basic guidelines for the trier of fact must be: (a) whether the average person, applying contemporary community standards would find that the work, taken as a whole, appeals to the prurient interest; (b) whether the work depicts or describes, in a patently offensive way, sexual conduct specifically defined by the applicable state law; and (c) whether the work, taken as a whole, lacks serious literary, artistic, political, or scientific value.

The third factor of the Miller test ensures that condition (i) is satisfied, and the first two factors ensure that condition (ii) is satisfied. If a work “lacks serious literary,
artistic, political, or scientific value,” then it presumptively has negligible social value. And if a work is “patently offensive,” adjudged by “contemporary community standards,” then it presumptively generates a substantial negative externality. Thus, works which satisfy the three factors are potentially “harmful speech” and may be subject to government regulation.

1.8 General Theories of Free Speech

Several other categories of speech have been the subject of proposals for speech regulation. For example, Professors Sunstein and Vermeule wrote an article on how the government might undermine the spread of conspiracy theories.\(^{128}\) Sunstein and Vermeule limit the scope of their inquiry to conspiracy theories which are “false, harmful, and unjustified,”\(^{129}\) breezing through consideration of their putative harmfulness.\(^{130}\) They proceed under the assumption that governments should seek to undermine the speech of conspiracy theorists when they are “false” (i.e., generate negligible value) and “harmful” (i.e., impose a negative externality). The question for them is not whether the government should suppress harmful speech, but rather what the best means of suppressing that speech might be. The idiosyncratic problem with conspiracy theories, by their lights, is their “self-sealing” and insular quality. Direct attempts at regulation are liable to backfire, drawing attention to—or even legitimizing—the very ideas the government seeks to suppress. Sunstein and Vermeule therefore suggest mechanisms which governments may use to “cognitively infiltrate” the spread of conspiracy theories.\(^{131}\)

Illustratively, Sunstein and Vermeule’s article elicited the very “backfiring” effect they feared direct regulation would create, enflaming conspiracy theorists who took the article as evidence of a conspiracy to suppress their ideas.\(^{132}\) This is surprising if only because one does not imagine conspiracy theorists regularly browsing the pages of The Journal of Political Philosophy. Regardless, for present purposes the takeaway is that Sunstein and Vermeule take it for granted that conspiracies theories ought not be protected speech, but for the special complications attending that phenomenon.\(^{133}\)


\(^{129}\) Id. at 207.

\(^{130}\) Id. at 220.

\(^{131}\) Id. at 218.

\(^{132}\) Andrew Marantz, How a Liberal Scholar of Conspiracy Theories Became the Subject of a Right-Wing Conspiracy Theory, The New Yorker, Dec. 27, 2017, at ??.

\(^{133}\) An earlier draft of the paper even advocated direct censorship in some cases. Cass Sunstein & Adrian Vermeule, Conspiracy Theories 20 (John M. Olin Program in Law and Economics Working
Similarly, Professor Lidsky undertook an analysis of Holocaust denial,\textsuperscript{134} arguing that Holocaust denial tends to cause a profound negative externality,\textsuperscript{135} while generating no value.\textsuperscript{136} Lidsky ultimately rejects direct regulation of speech for reasons not dissimilar to those Sunstein and Vermeule contemplate,\textsuperscript{137} however she also accepts the assumption that harmful speech ought to be regulated if such regulation can feasibly be implemented.

2 Refutation of the Harm-Based Argument

The foregoing section establishes that throughout the law and legal scholarship the \textit{harmfulness} of a category of speech is indeed adduced as a justification for its regulation. In this section, I present my counterargument to harm-based justifications. Recall the harm-deterrence argument is implied in support of premise (2):

2.1. If harmful speech is punished, then the supply of harmful speech decreases.

2.2. If the supply of harmful speech decreases, then social welfare is improved.

2. If harmful speech is punished, then social welfare is improved. (from 2.1, 2.2)

My objective is to show that sub-premises (2.1) and (2.2) are false.

2.1 Endogenizing the \textit{Harm} of “Harmful Speech”

Consider the inferential move in (2.2): that decreasing the supply of harmful speech will tend to effect an improvement in social welfare (or equivalently, a reduction in total harm). The assumption driving the received view is that the average magnitude of harm caused by any particular harmful expression is \textit{constant}, such that total harm varies with the quantity of harmful expressions. In other words, if one harmful


\textsuperscript{135}Id. at 1093 (“Both individuals and societies suffer harm as a result of Holocaust denial. Holocaust denial is a profound affront to human dignity. . . . If successful in gaining ‘converts,’ those who deny the Holocaust turn ‘history’ into an excuse for anti-Semitism and persecution.”)

\textsuperscript{136}Id. at 1095 (“That Holocaust denial is not only harmful but harmful and \textit{false} ought to create more leeway for punishing Holocaust Denial. The Holocaust is a historical fact, about as well documented as any historical fact could ever hope to be.”).

\textsuperscript{137}See id. at 1099.
expression generates social cost \( c \), and the supply of a harmful type of expression is \( \sigma \), then the total social cost is simply the harm multiplied by quantity, \( c \times \sigma \).

I contend that this formulation is too simplistic. In very many circumstances, the harmfulness of an expression will decrease as supply increases. In other words, the average per-unit harm \( h \) will be a function of the quantity of expressions \( \sigma \), such that marginal harm decreases as supply increases.\(^{138}\) Thus, in case the marginal harm varies with supply, the social cost will not be \( c \times \sigma \), but rather \( h(\sigma) \times \sigma \).

The question now arises whether this is a better characterization of harmful speech in fact. For the sake of organizational clarity, I postpone full discussion of this issue to §3 below. However, I should remark in passing the facial plausibility of this premise. Consider obscenity as an exemplar. The intuition is that increasing exposure to “obscene” speech will tend to render it commonplace and less scandalous. The phenomenon is easily observed in the divergence between American and European attitudes toward nudity.\(^{139}\) The relationship seems to be that the more common a putatively “obscene” expression is within a community, the weaker the negative reaction (if any). The intuition is easy to accept: that psychic phenomena like offense, credulity, outrage, or alarm obey the law of diminishing marginal (dis)utility. It is a lesson we all recognize from Aesop’s Boy Who Cried Wolf.\(^{140}\)

Now, if within a given interval \( \sigma \in I \), \( \sigma \) increases at a slower rate than \( h(\sigma) \) decreases,\(^{141}\) then it follows trivially that decreasing the supply of harmful expressions in that interval will have the effect of increasing total harm.

Consider the following numerical example. Suppose that \( h : \sigma \mapsto 400\sigma^{-2} \). This satisfies the relevant conditions.\(^{142}\) The following table, representing the values returned by the function, illustrates the point:

---

\(^{138}\)I.e., \( \frac{\partial h}{\partial \sigma} < 0 \).

\(^{139}\)See, e.g., Terrence Witkowskia & Joachim Kellner, Convergent, Contrasting, and Country-Specific Attitudes toward Television Advertising in Germany and the United States, 42 J. BUSN. Res. 167, 171, 172 (1998) (finding Americans were much less likely than Germans to regard television advertisements as depicting a “wholesome” world, despite occasional frontal nudity in German advertisements); Colin R. Harbke & Dana F. Lindemann, Acceptance of Female Public Toplessness: Structural, Contextual, and Individual Predictors of Support 27 CANADIAN J. HUM. SEXUALITY, 92, 97 (2018) (finding 18% of Americans absolutely opposed to female public toplessness, as compared with 0% of Europeans).

\(^{140}\)I.e., if \( \frac{\partial h}{\partial \sigma} \sigma < -h(\sigma) \) obtains. To see that this condition results in a decrease in social welfare, observe that social welfare is decreasing iff \( \frac{\partial}{\partial \sigma} (h(\sigma) \sigma) < 0 \). Therefore, \( \frac{\partial h}{\partial \sigma} \sigma + h(\sigma) < 0 \), which is equivalent to \( \frac{\partial h}{\partial \sigma} \sigma < -h(\sigma) \).

\(^{141}\)I.e., \( \frac{\partial h}{\partial \sigma} = -800\sigma^{-3} < 0 \) for all \( \sigma \in [0, \infty) \); and \( \frac{\partial h}{\partial \sigma} \sigma = -800\sigma^{-2} < -400\sigma^{-2} = -h(\sigma) \) for all \( \sigma \in [0, \infty) \).
Observe that although the supply of harmful speech $\sigma$ increases, the average per-unit harm $h(\sigma)$ decreases at a faster rate, and therefore total social harm $h(\sigma) \times \sigma$ will decrease as the supply of harmful speech increases.

Of course, the numerical example is not proposed to model any specific real-world class of expressive activity. The function $h : \sigma \mapsto 400\sigma^{-2}$ was chosen to illustrate a mathematical point. In some real-world contexts, the harm function may be constant or increasing, in which case sanctions on harmful speech would, ceteris paribus, effect an improvement in social welfare. But the point is that situations comparable to the numerical example are conceivable. And if such conditions are conceivable, then premise (2.2) is false, and the harm-deterrence argument is unsound.

As a practical corollary, even when $\frac{\partial h}{\partial \sigma} \sigma \geq -h(\sigma)$, so long as $\frac{\partial h}{\partial \sigma} < 0$, the marginal effectiveness of legal sanctions will be mitigated to some extent. In such cases, even though decreasing the supply of harmful speech would decrease the social cost arising from that speech, it may still be inefficient to sanction harmful speech when we factor in the costliness of detection, sanctions, and litigation costs.

To restate the foregoing discussion less formally, the intuition is this: the average potency of some types of harmful speech diminishes as instances of such speech become more commonplace. Inversely, the average potency of such expressions increases as instances become scarcer. It is possible therefore that the average potency of expressions increases faster than the quantity of expressions decreases, in which case total harm is greater when there are fewer instances of harmful speech. Granting “harmful expressions” are net harmful, the tradeoff is that fewer expressions entail greater per-unit harm, and more expressions entail less per-unit harm. The question is what the total effect is. It is conceivable that a decrease in the quantity of harmful speech actually increases the total harm arising from that category of speech by increasing per-unit harm. As shorthand, let us refer to this phenomenon—where decreasing supply increases potency—as the “potency effect.”

### 2.2 Endogenizing the Benefit of Harmful Speech

I next counter premise (2.1): the proposition that if harmful speech is punished, then the supply of harmful speech decreases. This premise may be further analyzed as relying upon the principle that the imposition of legal sanctions on certain kinds of
speech will tend to decrease the expected value of expression for prospective speakers of the relevant class. Assume that the prospective speaker of some harmful speech would derive some value from expressing it.

We may analyze the value the speaker derives from harmful speech into two components. First, the speech could generate value simply because the expression itself is gratifying to the speaker. Possibly it serves some innocent non-harm-directed purpose. Or possibly it is “therapeutic.” Second—and more interestingly—the speech could generate value because the harmful effect it causes is desirable to the speaker. In other words, the “harm” is not merely incidental, but the very object of the speaker’s intention. For convenience, let us call the first component “intrinsic value,” and the second component “extrinsic value.” I am principally concerned with the latter component.

Extrinsic value is a function of the magnitude of the externality. The greater its impact, the more value the speaker derives from engaging in it. This characterization is eminently plausible in speech contexts where the purpose of an expression is to shock, cause hurt, insult, or scandalize. In the category under consideration, extrinsic value $V_E$ increases as the average per-unit harm $h$ increases.\footnote{I.e., $\frac{\partial V_E}{\partial h} > 0$.}

The argument for the received view ignores the extrinsic value of harmful speech. Accounting for extrinsic value illuminates a second countervailing effect, which at a minimum weakens the argument for legal sanctions (and entirely nullifies it in the extremum case). For the purpose of illustration, suppose that the imposition of legal sanctions has the desired consequence of reducing the supply of some type of harmful speech. If the supply decreases, then assuming some level of potency effect,\footnote{I.e., $\frac{\partial h}{\partial S} < 0$.} the reduction in supply will effect an increase in the average per-unit harm of a given type of expression. This increases the extrinsic value of speech to prospective speakers, and thereby increases incentives to produce harmful speech, counteracting the initial reduction in supply.

Thus, in cases where potency effect and extrinsic valuation are present,\footnote{I.e., when $\frac{\partial h}{\partial S} < 0$ and $\frac{\partial V_E}{\partial h} > 0$.} increasing sanctions will have no effect whatever on the supply of

\begin{footnotesize}
\begin{enumerate}
\item[I.e.,] $\frac{\partial V_E}{\partial h} > 0$.
\item[I.e.,] $\frac{\partial h}{\partial S} < 0$.
\item[I.e., when] $\frac{\partial h}{\partial S} < 0$ and $\frac{\partial V_E}{\partial h} > 0$.
\item[Although, unlike] a physical spring, there is no particular reason to suppose the relation must be linear. Yet I am informed by my physics-degree wielding research assistant that not all springs respond linearly to force in any case.
\item[I.e.,] where $\frac{\partial V_E}{\partial h} \frac{\partial h}{\partial S} \frac{\partial S}{\partial S} > 1$, with $S$ denoting the sanction level and $\frac{\partial h}{\partial S} \leq 0$. In other words, if the effect of increasing sanctions decreases supply, and decreasing supply increases marginal harm, and
\end{enumerate}
\end{footnotesize}
harmful expressions.\textsuperscript{148} This defeats the claim that the imposition of legal sanctions on harmful speech necessarily decreases the incentives of prospective speakers to engage in it.

This general principle may also be observed in non-speech contexts. For example, in the market for elephant ivory, it has been argued that the imposition of legal sanctions on the international ivory trade, a policy adopted by 183 countries,\textsuperscript{149} has had the perverse effect of decreasing the supply of ivory, thereby increasing its market value, and thereby increasing incentives for poachers to hunt elephants.\textsuperscript{150} The underlying principle is essentially the same here.

Again, we have a practical corollary. In cases where the marginal effect of sanctions on the supply of harmful speech is \textit{de minimis}, the costliness of detection, sanctions, and litigation can tip the balance, such that the net effect could be a \textit{reduction} in social welfare when all factors are considered.

Restating the foregoing discussion less formally, the intuition is this: the speakers of some classes of harmful speech \textit{intend} their expressions to cause harm. The imposition of sanctions on that class of speech reduces the payoffs that speakers receive from expressing it, reducing the supply of harmful speech. However, if the reduction in the supply of such speech increases hearers’ sensitivity to that speech, they will be more susceptible to being harmed by it. This increases the benefit that speakers receive from expressing it, thereby increasing the supply of harmful speech. In the limiting case, the countervailing effect entirely cancels out the reduction in the supply of harmful speech. As shorthand, let us refer to this phenomenon as the “inverse supply effect.”

\subsection{2.3 Filtering}

For the third countervailing effect, I return to premise (2.2): that if the supply of harmful speech decreases, then social welfare is improved. Assuming the first two effects are present for some class of expressive activity, there arises the possibility of a selection effect, which further undermines the premise.

\begin{itemize}
\item \textsuperscript{148}I have in mind a formulation of the speaker’s utility function as being $U = V_I + V_E - S$, where $V_I$ denotes the intrinsic value of speech. This formulation ignores imperfect enforcement—without loss of generality.
\item \textsuperscript{150}For example, see Daniel W.S. Challender & Douglas C. MacMillan. “Poaching is More than an Enforcement Problem.” \textit{7 Conservation Letters}. 484–494 (2014).
\end{itemize}
If we assume that prospective speakers are heterogeneous in the magnitude of harm their speech causes, then the imposition of legal sanctions may not only affect the size of the population of speakers, but also the composition of that set. Let us partition the population of speakers into two subsets: high-harm speakers and low-harm speakers. The distinction is that high-harm speakers are those who—whether by circumstance or skill—are likely to cause greater harm than low-harm speakers when engaging in a type of speech. The construction is open to the possibility that an individual may be a high-harm speaker in one situation, but a low-harm speaker in another. For example, a given person who is artless in lying (and thus a low-harm speaker with respect to libel) may have a talent for causing offense (and thus be a high-harm speaker with respect to obscenity).

Now, suppose the conditions for the potency effect and the inverse supply effect are satisfied. It follows that low-harm speakers will be more susceptible to deterrence than high-harm speakers, because they derive less benefit from causing harm. This is definitional. A low-harm speaker is one whose harmful speech causes less harm than that of a high-harm speaker. Thus, the extrinsic benefit that the low-harm speaker derives must be less than the extrinsic benefit that the high-harm speaker derives, because the harm caused is less. Recall, extrinsic value is defined as increasing as the harm caused increases. Therefore, assuming the sanction level is held constant for all individuals, low-harm speakers will, ceteris paribus, be deterred more easily than high-harm speakers.

But if low-harm speakers elect to refrain from engaging in harmful speech, then they will have decreased the supply of total speakers, and by the potency effect, increased the marginal harm experienced by hearers. And by the inverse supply effect, this increases the incentives of speakers to engage in harmful speech. Specifically, it will tend to be high-harm speakers, less sensitive to the deterrent effect of sanctions, who are incentivized to “fill the gap” left by the low-harm speakers who were deterred.

Thus, even if the benefit of regulating speech outweighed the potency effect and inverse supply effect, it may still be the case that speech regulation would be inefficient. Although the supply of harmful speech may be reduced, nevertheless the altered composition of the population of speakers might result in an increase in total harm. The principle is intuitive and may be better communicated with the aid of a numerical example. The following table illustrates the point with hypothetical

---

151 We might formalize this by defining two distinct harm functions, such that the high-harm speaker’s expressions generate $h^H$ harm, and the low-harm speaker’s expressions generate $h^L$ harm, such that $h^H(\sigma) > h^L(\sigma)$ for all $\sigma$.

152 I.e., $\frac{\partial h}{\partial \sigma} < 0$ and $\frac{\partial V}{\partial h} > 0$. 

44
values. Assume that the average per-unit harm of low-harm speech is 5, and the average per-unit harm of high-harm speech is 10.

<table>
<thead>
<tr>
<th>Total Supply of Harmful Speech</th>
<th>Low-Harm Speech</th>
<th>High-Harm Speech</th>
<th>Total Harm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unregulated</td>
<td>10</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>Regulated</td>
<td>9</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

Note that these numbers were chosen not to represent a specific real-world effect, but rather to develop an intuition for the general principle. As shorthand, let us refer to this phenomenon—where the imposition of sanctions screens out low-harm individuals, increasing incentives for high-harm individuals to fill their place—as “filtering.” Again, in the interest of organizational clarity, I discuss the plausibility of filtering in real-world circumstances in §3 below.

To be sure, filtering is the most tenuous of the three countervailing effects. It is tenuous, firstly, because it assumes the presence of the other two effects. Whereas the potency effect may arise even when the conditions for the inverse supply effect and filtering are not satisfied, the inverse supply effect depends upon the presence of the potency effect. And filtering depends upon the presence of both the potency effect and the inverse supply effect. Secondly, filtering requires additional assumptions, which are disputable. It assumes that high-harm speech and low-harm speech of a given type are substitutes, differing only in magnitude. It also assumes that high-harm individuals are, apart from the definitional distinction, otherwise identical to low-harm individuals. However, it may be that low-harm individuals inflict low magnitude harm, because they differ in motivation. For example, it may be that high-harm individuals inflict high magnitude harm because their speech is principally motivated by a desire to cause harm, whereas low-harm individuals inflict harm only incidentally. If the reason why low-harm individuals inflict low magnitude harm is because they are motivated by other objectives, then the comparatively lower extrinsic value they derive (relative to high-harm speakers) may be offset by higher intrinsic value (i.e., their principal motive for engaging in such speech—whatever that might be). If this were the case, then my claim that low-harm individuals are more susceptible to deterrence would be mistaken. Nevertheless, I think filtering could arise in many speech contexts, and recognition of the possibility is sufficient to establish the falsity of premise (2.2).

And once again, there is the corollary that even if regulation were welfare-improving, despite the presence of all three countervailing effects, extending the model to account for the costliness of detection, sanctions and litigation can still render speech restrictive policies inefficient in practice.
2.4 Counter-Argument in General Form

In the preceding subsections, I have identified three countervailing effects, which undermine the argument for restrictions on harmful speech. The point is that merely identifying a class of expressive activity as satisfying conditions (i) and (ii) does not suffice to justify its regulation. Indeed, I identify conditions where regulation of harmful speech can even have the effect of increasing total harm.

To summarize: the potency effect identifies a tradeoff between the quantity of harmful speech and the average per-unit harm of that speech; the inverse supply effect identifies a friction which can undermine the effectiveness of sanctions on expressive activities, arising from increasing incentives to produce harmful speech when quantity decreases; and filtering identifies screening that could arise when speakers are heterogeneous in the magnitude of harm their expressive activity causes. Where speakers are analogized to producers and hearers are analogized to consumers, it may be helpful to think of the potency effect as a “demand-side” effect, and inverse supply and filtering as “supply-side” effects.

All three effects undermine steps in the argument for the received position. The conventional view is that imposing (or increasing) sanctions on harmful speech deters production of that speech, thereby effecting an improvement in social welfare. I demonstrate that the conclusion does not necessarily follow from these premises. Because premises (2.1) and (2.2) of the harm-deterrence argument can fail, it follows that premise (2) is not necessarily true, and therefore that the harm-based justification requires more than satisfaction of conditions (i) and (ii) to be sound.

It is important to distinguish what I am not saying. The presence of these effects does not establish that sanctions cannot be welfare-improving (or are always welfare-reducing). To assert this would be claim too much. Rather, what the possible presence of these effects does is it undermines an essential inferential component of the harm-based justification. It reveals two premises to be less sturdy than previously supposed. This raises the argumentative bar for the regulation of speech. It requires proponents of speech restrictions do more than merely point to the net harmfulness of a category of expression to justify regulating it.

It is also important to distinguish how my argument differs from other counterarguments opposing the regulation of harmful speech. Very many of the counterarguments opposing the regulation of harmful speech contest whether a putatively “harmful” expressive activity is truly harmful—when all factors are considered.153 These counterarguments seek to undermine claims that a category of expressive ac-

---

153 For example, arguments that unpatriotic dissent is desirable in a democratic society (contra efforts to suppress subversive speech). CITE Geoff Stone article.
activity satisfy conditions (i) or (ii). Although I am broadly sympathetic with this line of attack, this is not the nature of my counterargument here. My counterargument accepts, arguendo, the undesirability of putative “harms,” but contends that the imposition of sanctions may be ineffective at reducing that harm. Indeed, my argument suggests that sanctions may even exacerbate the harm in some circumstances.

Building on my negative argument, I have also hinted at a stronger, affirmative claim: that the presence of costly detection, sanctions, and litigation may tip the scale. *Even if* a speech regulation would reduce the direct harm from speech—despite the three countervailing effects—their presence would nevertheless mitigate the positive effects of such regulation. At this point, the enforcement costs (i.e., detection, sanction, and litigation costs), could push marginal regulation into the negative, so that it represented a social loss.

Yet aside from the enforcement cost “kicker,” I have another affirmative argument still. If the countervailing effects are sufficient to nudge the balance of social welfare close enough to the indifference point—where we are uncertain whether the total effect is positive or negative—then I contend the presumption ought to be against regulation, for if there is any meaning at all to the general speech right, it must at a minimum establish a default rule in cases of uncertainty. If it is not even that, then it is nothing at all. Ergo, when it is unclear whether a speech regulation would effect an improvement or reduction in social welfare, the presumption ought to be against government interference. This is not only what the law ought to be, but what the law is. The principle is implicit in nearly every judicial opinion: that in the absence of a harm-based justification, the government may not regulate speech.154

Thus, if the three countervailing effects succeed only in pushing the balance of social welfare to within the penumbra of uncertainty, two considerations will militate against the regulation of speech. First, the costliness of detection, sanctions, and litigation. Second, the default presumption of the general speech right against regulation.

### 2.5 Counter-Counterarguments Anticipated

I anticipate two potential objections. First, the critic might object that the three effects I have described heretofore are not unique to speech restrictions. I imagine such objections hitting upon the presence of the three countervailing effects in other realms of conduct, where a restrictive law is nevertheless deemed warranted. These objections would be in the class of *ad absurdum* arguments, possessing the following form:

---

154CITES
Some non-speech activity $X$ exhibits one or more of the three effects. When some of the three effects are present, regulation may be inefficient. Therefore, regulation of $X$ may be inefficient. Therefore $X$ should not be regulated absent further evidence.

But it is absurd to think that regulating $X$ could be inefficient. Therefore the inference—that regulation in the presence of the three countervailing effects might entail a reduction in social welfare—must be invalid.

The strength of this counter-counterargument depends entirely upon the strength of the premise that regulating $X$ is “obviously” efficient. Presumably, the proponent of speech regulation would choose an $X$ that most people would readily agree ought to be regulated.

For example, consider if $X$ were vandalism. The counter-counterargument would observe the possible presence of the potency effect, the inverse-supply effect, and filtering when regulating vandalism. Thus, if my argument were to be taken seriously, then it would follow that the law ought not prohibit vandalism. And yet, my hypothetical opponent would contend, it is obvious that the law should prohibit vandalism, and therefore the general principle of my argument fails.

I have three responses. First, assuming the wisdom of regulating $X$ is obvious, I would question whether one really could not satisfactorily provide the additional reasons which justify the inference to premise (2). In the case of vandalism, I think it may be plausibly argued that decreasing the quantity of vandalism would not necessarily increase the average per-unit harm of the undeterred residual vandalism. In this case, it may simply be observed that the potency effect is not present, and since the inverse-supply effect and filtering depend upon the potency effect, we may conclude that none of the three countervailing effects obtain in fact.

However, if it were established empirically that the potency effect were present for some $X$, then my second response would be to inquire whether there exist kickers militating against regulation of $X$. Again, in the case of vandalism, there does not seem to be any general category of behavior, of which vandalism is a subset, which the law accords as a general right. Also, it seems that enforcement costs may be offset by the cost of repairing the vandalized property. Therefore, even if it were uncertain what effect vandalism regulations would have on social welfare, there would be no affirmative reason not to regulate it.

And finally, if it were established that the potency effect were present for some $X$, and if there were a general right to some broader class of activities which included $X$, then my response would be to embrace the “absurdum” conclusion. Supposing it were shown that vandalism regulations exhibited the three countervailing effects, and
supposing there were some general right that vandalism regulations would incur upon, I would incline to accept the conclusion that vandalism ought not be prohibited. I am skeptical of course that either of the conditions could be met in the case of vandalism particularly, but if it could be shown that the three countervailing effects were truly present, and the balance of social costs were unclear, and there existed kickers which tended to militate against regulation, I should think it not at all “obvious” that X should be regulated.

The second objection I anticipate is more general. Some readers may complain that my negative argument proceeds from armchair observations about a theoretical model unsupported by empirical data. The hostile critic will protest that I do not even attempt to determine the values of variables nor to discern concrete facts.

This grossly misses the point. As I demonstrate in §1, the conventional justification for speech regulation relies upon an economic inference—the harm-deterrence argument. The harm-deterrence argument depends upon premises which are theoretical economic claims. The counterargument which I provide is correspondingly a theoretical economic counterargument.

Complaining that my counterargument fails to delve into empirics mistakes the logical structure of my contention. To illustrate the point, suppose a person asserted the argument: all cats are quadrupeds; all dogs are quadrupeds; therefore all cats are dogs. This is clearly a textbook “illicit minor” fallacy. The proper counter here is to demonstrate the invalidity of the inference: that “all X are Y” and “all Z are Y” does not entail “all X are Z.” The counter requires no empirical investigation into the nature of cats, dogs, quadrupeds, or any other possible X, Y, or Z.

Similarly, the received justification for speech regulation relies upon an economic inference. The presence of the three countervailing effects undermines that inference. No empirics are necessary to demonstrate the unsoundness of the inference. Indeed, empirical claims are entirely irrelevant to the point in dispute. To insist otherwise is to misunderstand the nature of scientific theory generally.

3 Applications

I might reasonably end this article here. My objective being to show that the harm-deterrence argument is unsound, it suffices to demonstrate the possible occurrence of the three countervailing effects to establish my thesis. However, some intuitive connection to real world controversies may be wanted. I expect both proponents and opponents of speech regulation will be interested not only in whether the three countervailing effects are possible, but whether they are plausible. In other words,
whether there is a coherent intuitive story where the three countervailing effects are believably present in areas of controversy.

In this section, I consider several prominent examples, supplying reasons to believe that potency, inverse supply effect, and filtering might arise. I do not intend these examples to be exhaustive. My counterargument is deliberately framed in general terms, and I expect the three countervailing effects might arise in any circumstance where harm-based justifications for regulation are proposed. Neither do I intend the examples to be conclusive. A rigorous empirical investigation into any of the particulars would require considerably more than I could hope to furnish within the scope of this article. My aim in this section is simply to sketch out the facial plausibility of the effects obtaining in a variety of contexts.

3.1 Affronting Speech

Let us start by considering a subset of harmful speech, the supposed harm of which is psychic injury. Let us refer to this subset as “affronting speech.” Affronting speech represents the simplest and most straightforward context to which my argument applies.

There are two relevant player categories in the context of affronting speech: injurers and victims. The conventional justification for the regulation of affronting speech is that injurers impose negative externalities upon victims, reducing social welfare. Regulation is meant to deter injurers from expressing harmful speech. To reiterate: I take the characterization—that the relevant harm is the psychological effect it has upon victims—as defining the subset of speech which is affronting.

Affronting speech is analyzable into familiar doctrinal types. I include in the category of “affronting speech”: offense, fighting words, hate speech, obscenity, indecency, and profanity.

3.1.1 Offense

In the context of offensive speech, the relevant “affront” is the displeasure experienced by some portion of the public when a cherished symbol or idea is debased. Typical examples include the desecration of flags or religious symbols.155

---

155 An especially broad curtailing of speech rights on the basis of offense may be observed in the jurisprudence of Poland. Curiously, the Polish Supreme Court has ruled that the freedom of religion includes the right of religious people “not to be offended.”

“The subject of the protection to which Article 196 of the Criminal Code pertains are the religious feelings arising from the constitutional freedom of religion. It is accepted in the jurisprudence that the religious feelings are, due to their nature and the direct
Here, the presence of all three countervailing effects may be observed. First, it is surely the case that marginal harm decreases as quantity increases. To take the most notable exemplar of “offensive speech,” observe that public sentiment toward the proposal for a constitutional prohibition on flag-burning has, since Texas v. Johnson in 1989, steadily waned. Without intending to diminish the possible contribution of other causal factors, I would nevertheless expect the magnitude of popular support for a “Flag-Burning Amendment” to be a reasonable proxy for the magnitude psychic harm that the public experiences when witnessing immolation of the flag.

This is of course a thoroughly unsurprising fact. It cannot reasonably be doubted whether an expression which elicits hostility will tend to diminish in potency as individuals susceptible to offense grow inured to the stimulus through repeated exposure.

link with freedom of religion, subject to special protection. . . . There is no doubt that the protection of religious feelings, and thus human emotion associated with the faith professed by the individual, is also linked to the protection of the inherent and inalienable dignity of the human person, which is the source of freedom and human and civil rights [citing Article 30 of the Polish Constitution] . . . Article 196 of the CC serves as an expression of a specific position taken by the legislature in the potential conflict of freedom of expression . . . The legislature decided the conflict in favor of freedom of religion, assuming that it is unlawful to express the views that consist of insulting of religious object or place intended for public performance of religious ceremonies, which leads to offending religious feelings of others.”

Judgment of the Constitutional Tribunal of October 6, 2015 (Case no. SK 54/13).


For example, I imagine that the expressive effect of the law would be pronounced in these circumstances. CITE McAdams paper and book. The expressive factor—although not identical to the potency effect—would tend to reinforce my arguments regardless.
Common sense and everyday experience confirm this phenomenon.\footnote{CITE PSYCH LITERATURE}

Second, the inverse-supply effect requires that the benefit to injurers increase as harm increases. This is plausible for offensive speech. Returning to our exemplar, it is surely the case that those who burn flags do so precisely because they anticipate the offense it will cause. Indeed, it is difficult to imagine any other reason to endure the expense, risk of injury, and noxious fumes which the act entails, except to agitate the passions of those spectators who witness the incineration of their beloved symbol.

Unfortunately, there do not exist good statistics on the frequency of flag-burning. Yet even if there were, the other relevant variable—i.e., the magnitude of outrage—would remain difficult to quantify. It is telling however that the incidence of flag-burnings seems to increase precisely during those periods when proposals for flag-burning prohibitions are prominent in public discourse.\footnote{CITE BARLETT’S} The economic explanation is obvious: proposing to ban the burning of flags is tantamount to a declaration that one is acutely susceptible to its intended effect. It signals to injurers a high “price” they can extract from offensive expression, which naturally elicits an increase in production.

Additionally, the inverse supply effect also has an interesting interplay with publicity. News outlets, motivated to attract a novelty-craving audience, have greater incentivizes to report on seemingly uncommon events. Thus, even if the imposition of sanctions were to reduce the supply of injurers, this may have the perverse effect of amplifying instances of a type of speech so as to affect more recipients. As the publisher Alfred Harmsworth once famously quipped, “When a dog bites a man, that is not news, because it happens so often. But if a man bites a dog, that is news.”\footnote{CITE BARLETT’S} Correspondingly, when the “news” is something many people find offensive, the consequence of decreasing supply is increasing publicity, and with increasing publicity, increasing harm. The publicity effect can be regarded as a species of (or at least approximately equivalent to) the inverse supply effect.

Third, it seems plausible that filtering would also arise in the context of offense. Again, to take flag-burning as an example, we expect that the imposition of sanctions would naturally increase the expected cost of expression. Given that the inverse-supply effect obtains, injurers will tend to derive greater utility the more widely publicized their flag-burning, and less utility the less widely publicized. Thus,\footnote{For example, after President-elect Trump suggested that flag-burners should suffer criminal punishment (or a loss of citizenship), a small band of protestors gathered in front of Trump’s New York Hotel to burn American flags. Stapleton, Shannon. “Trump flag-burning tweet leads activists to burn some flags in New York.” Available at: https://www.reuters.com/article/us-usa-trump-flag/trump-flag-burning-tweet-leads-activists-to-burn-some-flags-in-new-york-idUSKBN13P06L.}

\footnote{CITE BARLETT’S}
injurers will be deterred from burning the flag in circumstances where the harm is low (i.e., when there are fewer potential recipients/victims), reducing supply, and increasing the potency of the expression. The effect is that the grandiose act of flag-burning will be reserved for those occasions when it will have the most impact. The savvy protester will refrain from burning the flag when few passersby will notice, avoiding the disutility of sanction. He will reserve his shock tactic until he has attracted the attention of photographers and journalists—to maximize the value he derives from incurring the sanction.

It is thus eminently plausible that all three effects would arise in the regulation of flag-burning. And the intuitive story maps easily to offensive speech generally. Whether the vehicle for expression takes the form of burning flags, draft cards, or books, carrying offensive signage, or printing offensive slogans on tee-shirts, we should expect increasing the supply of such expressions will tend to reduce their impact. And it is a safe surmise to suppose the speakers derive utility from the disutility they cause in others. And increasing the cost of expression through sanction would tend to deter low-harm instances more than high-harm instances, as strategic speakers “reserve” their expression for those occasions when it is likely to have the most impact.

3.1.2 Obscenity, Indecency, and Profanity

Similarly, in the context of obscenity, indecency, and profanity, we may conceive the relevant affront to be the displeasure experienced by some portion of the public when some “vulgar” expression is given voice. Typical examples include the portrayal of sexual activity, defecation, or urination; utterances of “dirty words”; depictions of nudity; devices and implements of sexual fetishists; and media tending to the subversion of social norms.\textsuperscript{162}

Let us take the “shock art” movement as our first exemplar.\textsuperscript{163} A cursory investigation of its history reveals robust examples of the potency effect. Among the principal developmental mechanisms of “shock art” seems to be simple one-upmanship. Each subsequent generation of shock artists seeks to unsettle conventions more rad-

\textsuperscript{162}Curiously, this has, in many totalitarian regimes, included abstract artistic works with little or no representational content whatever. For example, [DISCUSS Entartete Kunst in NAZI GERMANY]. It is ironic that historically the most ferocious attacks on art have targeted those works which have possessed the least political content.

\textsuperscript{163}Notable representatives of this “movement” include Andres Serrano, who photographed a crucifix submerged in his urine (\textit{Piss Christ}, 1987), Chris Ofili, whose \textit{Holy Virgin Mary} (1996) depicts the eponymous subject as a Black woman, with one exposed breast constructed from varnished elephant dung, against a collage of pornographic imagery, and Rick Gibson, who fashioned earrings from freeze-dried human fetuses (\textit{Human Earrings}, 1987).
ically than their predecessors had done. This may not be the only mechanism at work, however it is surely one of the main factors driving the genre.

The phenomenon follows from the principle: that scandals subside. It is the artist’s analogue to the second law of thermodynamics. Entropy, it seems, can be discovered in many forms. Marcel DuChamp’s once outrage-inducing *Fountain* (1917) sits today in the Museum of Modern Art in New York, evoking little reaction but the cool insouciance of perplexed tourists. The premiere of Stravinsky’s *Rite of Spring* in 1913 induced a riot. It is now a staple of the orchestral repertoire; and its performance provokes little more than dumb indifference from geriatric concertgoers struggling with their hearing aids.\(^{164}\) What is an artistic affront today is merely banal tomorrow. Its potency diminishes as audiences grow inured to its novelty.

Evidence of the inverse-supply effect is no less difficult to discern in the context of the arts. Discounting the attestations of shock artists themselves (their propensity to use public statements to subvert expectations renders their putative self-reflections incredible), it certainly seems from observation of their behavior that they are deliberate in their attempts to elicit “shock” responses.

This is wholly consistent with economic principles. The shrewd artist, seeking novelty, attention, notoriety, and wealth is wise to court controversy. Thus, reducing the quantity of shock art would tend to increase incentives for shock artists to create it. In other words, if the imposition of sanctions on obscene art reduced the supply of shock art, then by the potency effect it would increase audiences’ sensitivity to shock, thereby increasing the incentives for shock artists to produce it, and by the inverse supply effect counteract the initial reduction in supply.

The conditions for filtering seem to be satisfied also. Low-harm artists, whose work might incidentally cause shock, or who lack talent for producing shocking art, would be more susceptible to deterrence, because they derive less benefit from their relatively less effective attempts to elicit outrage. If they exit the market, decreasing the supply of obscene art, then by the potency effect, audiences’ sensitivity to obscenity will increase. And those artists whose comparative advantage lies in their talent for shocking will experience stronger incentives to produce obscene art.

All three effects are therefore plausibly present in the context of shock art. It follows that imposing sanctions on shock art could result in a net reduction in social welfare, undermining harm-based justifications for censorship of it. The critic may object that I am cherry-picking in my choice of exemplar—that I have deliberately

\(^{164}\)I do not mean to imply that the *Rite of Spring* was conceived as “shock art.” Arguably, no analogous movement ever arose in music—although George Antheil seems to have invested some effort in assuring that the premiere of *Ballet Mécanique* (1926) resulted in scandal, and of course John Cage also routinely courted controversy.
chosen shock art, because it is an especially availing context in which to observe the three countervailing effects. This misses the point. Shock art is acutely relevant, because it is the most liable sort of artistic expression to be regulated. It is the frontline in the battle for free speech. I am happy to concede that the inverse supply effect is unlikely to arise if governments were to regulate, for example, landscape paintings or muzak. This is irrelevant. The implausibility of the countervailing effects arising from the regulation of landscapes and muzak follows directly from the implausibility of their causing harm in the first place.

A still more vivid illustration of the three countervailing effects may be observed in the regulation of “pornographic” films. The earliest cinematic depiction of a remotely sexual nature is The Kiss (1896). One of the first commercially distributed films, The Kiss depicts a fully clothed man and woman nuzzling and exchanging brief pecks over the course of eighteen seconds (looped thrice). The Kiss was distributed by Thomas Edison in an effort to promote the kinetoscope—a nickel-operated, hand-cranked motion picture invention.\textsuperscript{165}

The kiss depicted in The Kiss is not, by modern standards, a remarkably passionate osculation. Yet it was deemed reprehensible in the twilight of the nineteenth century, provoking one critic to write:

\begin{quote}
    The spectacle of the prolonged pasturing on each other’s lips was hard to bear. When only life size it was pronouncedly beastly. Magnified to Gargantuan proportions and repeated three times over, it is absolutely disgusting.\textsuperscript{166}
\end{quote}

The Catholic Church denounced The Kiss, calling for censorship.\textsuperscript{167} Outrage over the film had hardly subsided before Edison released a still more scandalous vignette, Dolorita’s Passion Dance (1897), depicting a (once again fully clothed) woman engaging in an Iberian “passion dance.” So intolerable were her bodily contortions it seems that the sage authorities of New Jersey were left with little choice but to raid the Atlantic City parlor, in which it was being shown, and to ban the film.\textsuperscript{168}

The modern media consumer need only introspect, to seek out even a quantum of revulsion in his own mind upon viewing The Kiss, to recognize the potency effect at work. Indeed, I doubt modern viewers would recognize The Kiss as being erotic in

\textsuperscript{165}See generally Geltzer, Jeremy, Dirty Words and Filthy Pictures: Film and the First Amendment (2016) for a thoughtful historical survey of censorship in cinema.

\textsuperscript{166}Id.

\textsuperscript{167}Id.

\textsuperscript{168}Dolorita’s Passion Dance was thus the first censored work of cinema. It was also, at the time the mayor’s order was issued, the “most viewed Kinetograph picture the parlor had ever hosted.” Geltzer, 9, citing.
any sense at all. Whatever harms contemporary moral figures may have experienced in the furor over *The Kiss* are wholly incomprehensible in the present day, and the complete dissipation of that harm is undoubtedly due to the overwhelming quantity of onscreen kissing in television and cinema.

Evidence of the inverse supply effect can also be observed in the early days of cinema. The notoriety of *The Kiss* inspired imitators, presumably seeking to exploit the titillating potential of the new medium. Yet the erotic novelty of *The Kiss* and its copycats abated rapidly. While the demise of the kiss porn genre was likely due to the emergence of more extreme content, the requisite principle is retrievable: that injurers derived a benefit positively correlated with the magnitude of offense caused. As the sensitivity of audiences to onscreen kissing diminished, so too did incentives to exploit that effect.

It is true that onscreen kissing remains exceedingly common in films today, and the inverse supply effect implies that as audiences grow inured to the effect, incentives to depict it should decrease. Yet it is important to distinguish that the cinematic depiction of kissing was not exclusively or (except perhaps in the very earliest days of cinema history) even predominantly motivated by the pursuit of extrinsic value. The portrayal of kissing has substantial intrinsic value. It is a useful expository tool in the construction of stories—in depicting the relationships between characters, and revealing characters' emotional states visually. The speaker's extrinsic motivation—the benefit received from causing offense—in the depiction of kissing seems entirely absent in the modern filmmaker.

Likewise, filtering also seems to be present. Whereas mainstream filmmakers, for whom eroticism was but an incidental element of their art, sought to stay within (if only just within) the boundaries of what was permissible, the scofflaws whose very purpose was to produce lascivious content would exploit the regulatory effect of censorship, charging high prices for content which prevailing social norms would ensure possessed a high degree of potency.

Now, proponents of obscenity, indecency, and profanity regulation may counter that I have missed the point entirely. They will not dispute that the effect of pornographic imagery tends to dissipate with increased exposure. Indeed, this is the

---

169 *The Kiss in the Tunnel* (1899) and *The Kiss* (1900).
170 For example, Georges Melies' *Après Le Bal* (1897, Fr.), depicting a fully nude woman, and *A L'Ecu d'Or ou la Bonne Auberge* (1908, Fr.), depicting penetrative intercourse.
171 The mechanism has one complication. The pornographer seems not to have been motivated to *cause* offense directly. There is an intermediate inference. Content which is sufficiently novel to stimulate erotic sentiments for a segment of the population will tend to be sufficiently novel to offend the indignation of the other. The effect will follow, even if the extrinsic valuation is not direct.
gravamen of their complaint. They will concede that liberal speech laws dilute the harm of *The Kiss*, yet they will insist that a society indifferent to “corrupting imagery” is somehow a *worse society*. The *harm*, they might argue, is not the affront, but rather the coarsening of standards which preempts affront. Their argument is that it would be as though a patient went to a doctor complaining of soreness in his arm, and the doctor “cured” the ailment by severing a nerve so that the patient lost all feeling that limb.

This counter does not touch my argument. Indeed, it concedes to it. The argument that a society more sensitive to harms is somehow a “better society” relies on premises which cannot easily be reconciled with a welfarist model. It is a moral argument masquerading as a consequentialist argument. And indeed I think its irreconcilability with a welfarist model is itself evidence that there is no substance whatever in the claim.

Nevertheless, I think a brief digression may be warranted on this flimsy contention. I have two responses. First, to the extent that proponents of the obscenity, indecency, and profanity exception would claim that censorship is something like a defense of a society’s moral identity, they are taking up the losing side of a settled controversy. They revisit the arguments of Lord Patrick Devlin in his debate with H.L.A. Hart, in which Hart conclusively prevailed. Second, it seems doubtful that any person today would honestly contend that society was somehow better off (even “morally” better off) when a grainy close-up shot of a poorly-aimed kiss constituted a scandal. Can any proponent of the obscenity, indecency, and profanity exception really believe that our present indifference to *The Kiss* is a detriment? The question may be extended to later targets of censorship. Would our society be better off without films like *Scarface* (1932) or *Monty Python’s Life of Brian* (1979)? Would we be better off if Lenny Bruce or George Carlin were deterred from the expression of profane comedy? Would we be better off if the censors had prevailed in extinguishing James Joyce’s *Ulysses* or Nabokov’s *Lolita* from bookstores? I expect no reasonable person would answer these questions in the affirmative.

The modern proponent of the obscenity, indecency, and profanity exceptions may concede that these historical examples were unwarranted government intrusions, and that our society is no worse off for tolerating these former targets of speech regulation after all. But he may contend *those* obscenities were never *really* obscene in the first place. He will maintain that the obscenities, indecencies and profanities he perceives today are *different*. These things *really* are obscene, indecent, or profane. And *this time*—he will insist while marveling at the stupidity of his censorious predecessors

---

172I have nothing to contribute to Hart’s refutation, which I consider conclusive. Readers interested in these arguments are encouraged to consult the primary sources. CITE.
in drawing the line so poorly—this time he has surely gotten it right.

There is no scientific answer which would satisfy the moral ideologue, though I am skeptical whether he deserves a response at all. Regardless, for present purposes it suffices to observe the plausibility of the three countervailing effects arising in the context of obscenity, indecency, and profanity regulation.

### 3.1.3 Hate Speech

Hate speech is another context in which the three countervailing effects are very likely to arise. Notwithstanding the Supreme Court’s protean positions on the controversy, it seems quite clear that hate speech does possess negligible value and imposes a high social cost.

The hate speech context provides a unique opportunity for observation, because the social sanctions attached to bigoted expressions are especially severe. And the absence of legal sanctions (in the United States) tends to reduce speakers’ efforts at concealment. The presence of sanctions and absence of centralized enforcement allow us to easily observe how individuals tend to respond to the imposition of informal speech regulation.

Let us first consider the potency effect. One can hardly fail to observe that the psychic harm arising from racist, sexist, homophobic, or ableist expressions exhibits an inverse relationship to their frequency of use. For example, utterance of the term “nigger” has been abolished from civilized discourse. It is never uttered, but only referenced as “the N-word.” Due to the infrequency of its expression, the term is imbued with great weight. Its mere utterance, when directed toward a Black

---

173 In the United States, hate speech has enjoyed relatively strong Constitutional protection. Other putatively liberal democracies have taken a less tolerant stance toward hate speech. See, e.g., §130 of the German criminal code, Strafgesetzbuch, StGB promulgated on 13 November 1998 (Federal Law Gazette I, p. 945, p. 3322) (“(1) Whoever, in a manner that is capable of disturbing the public peace: 1. incites hatred against segments of the population or calls for violent or arbitrary measures against them; or 2. assaults the human dignity of others by insulting, maliciously maligning, or defaming segments of the population, shall be punished with imprisonment from three months to five years.”), the French penal code, Code Pénal R. 624-3-4 (prohibiting non-public defamation or insults to individuals on the basis of ethnicity, nationality, race, religion, sex, or sexual orientation), and in the United Kingdom, The Public Order Act 1986 (c 64) §4A, amended by §154 of the Criminal Justice and Public Order Act 1994 (“A person is guilty of an offence if, with intent to cause a person harassment, alarm or distress, he (a) uses threatening, abusive or insulting words or behaviour, or disorderly behaviour, or (b) displays any writing, sign or other visible representation which is threatening, abusive or insulting, thereby causing that or another person harassment, alarm or distress.”)

174 See supra §1.4.
individual, rivals the harm of physical violence. When uttered among non-Black individuals, it taken as conclusive evidence of a heinous defect in the speaker.

There can be no doubt as to the loathsomeness of racial bigotry and its effects, and the desire to do something to curb its harms is understandable. Yet we should be cognizant of the consequence of reducing the incidence of expressions of hate. Eradicating racial epithets from civilized discourse magnifies their power. It hands to the unapologetic racist, unencumbered by civilized norms, a potent weapon.

This insight is not novel, of course. Discriminated communities have seemingly intuited the danger of the potency effect which arises from the policing of language. Black comedians and musicians have for many decades incorporated use of the term “nigger” into their routines and lyrics. In much the way a vaccine immunizes patients to disease, repeated utterance of the term in innocuous contexts mitigates its potential to hurt in more virulent circumstances. The phenomenon, which activists have termed “reappropriation” or “reclamation,” has attracted some scholarly attention, although the research (mainly in sociology and cultural studies departments) wants somewhat for rigor. Analysis in terms of the potency effect represents a possible avenue for improvement.

Reappropriation is a widespread strategy. Although its use by Black Americans furnishes a prominent and easily recognizable example, we can observe a variety of discriminated groups employing it. The disabled community has reappropriated the term “cripple.” The Asian American community has attempted to reappropriate “slant.” And the gay community seems to have been especially successful in diffusing the pejorative connotations of words like “queer” and “dyke.”

Reappropriation is not a recent phenomenon. Deliberate attempts to exploit the potency effect have occurred throughout history. For example, in the eighteenth century, Protestant followers of John and Charles Wesley were pejoratively labeled “methodists.” The group embraced that term, accepting it as the proper name of their denomination. So successful was the reappropriation, few people are even aware that the term was once used to denigrate the adherents of that faith.

Let us next consider the inverse supply effect. It follows trivially from the extrinsic value that bigoted speakers derive from causing hurt that the conditions for the inverse supply effect will obtain in the context of hate speech. Given that there is a

---

175 More serious scholarship investigating the causes of semantic change exist of course in the study of linguistics. See W.V. Quine, *Quiddities* 53–54 (Belknap Press 1987).

176 Interestingly, a rock group consisting of Asian members, calling themselves “the Slants,” attempted to register for trademark protection, which the Patents and Trademark Office rejected, essentially on the ground that expressions of hate speech would not be granted trademark protection. The case was litigated to the Supreme Court, where musicians prevailed. See Matal v. Tam, 582 U.S. ____ (2017).
potency effect—i.e., that decreasing the supply of hateful expressions increases hearers’ sensitivity to them—individuals wishing to cause harm will experience stronger incentives to exploit that increased sensitivity. Reducing the supply of hate speech will thus tend to increase incentives for bigots to express it.

Lastly, let us consider filtering. Filtering is the most evident of the three countervailing effects in the context of hate speech. It is practically axiomatic that less racist individuals are more easily deterred from expressing hate speech, and more racist individuals are less easily deterred. There will also be marginal individuals, who are deterred from engaging in hate speech day-to-day, but who deploy the language only on those occasions when they feel it will have the most impact. The increased potency and asymmetric deterrence of hate speech regulation will tend to filter out the low-harm expressions and increase incentives for high-harm expressions.

3.1.4 Fighting Words

Another context in which the three countervailing effects arise is in fighting words. The analysis requires some finesse. The justification for regulating fighting words lies not in the harm caused by speech, but rather to the harm which results from the harm caused by speech. The concern is not the psychic injury done to the hearer of fighting words, but rather the consequent public disorder that results when the hearer seeks reprisal through violence. In other words, the primary harm (affront to the hearer) is the catalyst for secondary consequential harms (the public disorder and injury resulting from reprisal). The policy objective is to reduce the secondary harm, and mitigating the magnitude of primary harm—it is supposed—will tend to have the knock-on effect of reducing the secondary harm.

However, regulation of the primary harm is susceptible to the three countervailing effects. Consequently, it is conceivable that regulation would reduce rather than improve social welfare. Consider that if the prevalence of fighting words increases, hearers’ sensitivity to the affront will tend to decrease, reducing the probability of violent reprisal. Conversely, if the prevalence of fighting words decreases, then hearers’ sensitivity to the affront will tend to increase, raising the probability of violent reprisal. The concern with the knock-on effect does not alter the potency effect analysis.

Likewise, the secondary effects objective, idiosyncratic to fighting words, amplifies the inverse supply effect and filtering. Consider that low-harm injurers are more likely to be deterred by regulation (than high-harm injurers), because their efforts to cause affront are less effective. And high-harm injurers will thus experience increasing incentives to produce fighting words. High-harm injurers may be comparatively more
effective at producing fighting words, either because they have a talent for invective, or because are especially good at choosing targets more susceptible to injury.

Moreover, victims are also filtered, because hearers insensitive to fighting words will tend not to find themselves in a circumstance requiring litigation. This leaves only injurers who are especially skilled at insult, and hearers who are especially liable to violent retaliation.

The narrative is a plausible one. A community, where insults and threats are commonplace, is one where disparaging remarks are more likely to be shrugged off or ignored. An individual accustomed to coarse conversation is less likely to feel the cut of an affronting jibe, and less likely to feel inspired to respond with violence. It is the genteel victim, unfamiliar with harsh treatment, who is liable to feel an obligation to vengeance. So far as the objective is to reduce the incidence of public disorder, the better strategy is to encourage policies which inure the population to affront. The alternative is to cultivate a community of eggshells, astounded by insults to their honor, and inspired to dueling at the slightest provocation.

3.2 Persuasive Speech

Let us turn now to the regulation of harms which are supposed to result when hearers are convinced to adopt false beliefs or odious preferences. I will refer to expressions intended to affect hearers’ beliefs and preferences as “persuasive speech.” In contrast to affronting speech, the harm is not to cause offense, outrage, hurt, or distress in the recipient. Rather, the harm of persuasive speech is in the proliferation “bad ideas.”

I include within this category several subcategories, including defamation, commercial speech, and “fake news.”

3.2.1 Defamation

Defamation can occur when a group or individual transmits a signal to other individuals about a third party. There are four relevant player categories in the defamation context: (1) the purveyors of true information, (2) the purveyors of false information, (3) the recipients of information, and (4) the subjects of information.

There are two economic justifications for the enforcement of defamation claims. First, that defamatory speech generates an externality. Purveyors of false information enjoy some utility by sending false signals about subjects. And subjects suffer harm to their reputations. Assuming the value (to the purveyor) of transmitting false signals is less than the harm caused (to the subject), the conventional view is that the law should deter the sending of false signals. The argument is that by imposing sanctions on purveyors of false information, the expected benefit (to the purveyor) of
purveying false information decreases, and therefore the activity level of purveying false information decreases, effecting a reduction in social cost.

The second justification is that defamation laws have a screening effect. Recipients enjoy utility from the receipt of true information and disutility from the receipt of false information. By imposing a sanction on purveyors of false information (assuming purveyors of false information experience a higher probability of sanction than purveyors of true information), the asymmetric deterrence causes the supply of false information to decrease faster than the supply of true information (at least up to some socially efficient point), leading to a net improvement in social welfare. In other words, enforcement satisfies the monotone likelihood ratio property, increasing the ratio of purveyors of true information relative to purveyors of false information.

Notice that both justifications depend upon the belief level of recipients. If recipients of false information disbelieve, then neither the recipients nor the subjects suffer any harm. It is only when recipients of false information believe the false information that they and the subjects suffer harm. Thus, the problem may be simplified to some extent by focusing our analysis on the belief level of the recipients.

All three countervailing effects are present in the defamation context. Consider the harms suffered by recipients of false information. Observe that recipients' welfare will tend to increase, the greater their belief level in false information, and the weaker their belief level in true information. In other words, if some information $X$ is expressed, and a recipient’s belief level is represented by $\alpha \in [0, 1]$, where 0 is complete disbelief, and 1 is complete belief, then the recipient’s welfare function $U_R$ will be decreasing as $\alpha$ increases ($\frac{\partial U_R}{\partial \alpha} < 0$) if $X$ is false, and increasing as $\alpha$ increases ($\frac{\partial U_R}{\partial \alpha} > 0$) if $X$ is true. The trouble, of course, is that with any given signal $X$, recipients will not know whether $X$ is true or false without further investment in search.

If we assume that recipients are rational, and their beliefs are Bayesian, then they will set their belief level according to the prior probability that $X$ is true, and update their belief level to account for the probability that the truth of $X$ produces a confirming signal. In the simplest case, where recipients have no additional information about the relative credibility of a particular purveyor, they will take the probability that a given signal is true to be the supply of true signals $\sigma_T$ divided by the total supply of signals $\frac{\sigma_T}{\sigma_T + \sigma_F}$ (where $\sigma_F$ is the supply of false signals). It follows immediately that if the imposition of sanctions decreases the supply of false speech relative to true speech, then the belief level for any given expression will increase. And clearly, there will exist values for which the decrease in the supply of false speech is more than offset by the harm due to the increased belief level.

The intuition here is straightforward. Imagine a world in which everyone defames
and no one tells the truth. Clearly, the rational recipient would disbelieve all signals, and the harm caused by false signals would be zero. Now, as the proportion of truth-tellers increases, the credibility of a signal will tend to increase, and the marginal increase in harm for any given instance of defamation will tend to increase as well. Thus, increasing the supply of truth-tellers relative to the supply of defamers will tend to increase the per-unit harm of defamation. In other words, a recipient will tend to rely more when the ratio of purveyors of true information relative to purveyors of false information is higher.

However, defamation represents a special case. Although the potency effect is clearly present, in the defamation context it cannot result in a net increase in harm. If the increasing belief level generated a decrease in the recipient’s payoffs, then the rational recipient will behave as if the belief level were less than their actual belief level. Framed differently, we might disambiguate belief level from “trust level,” where the trust level $\beta \in [0,1]$ is a function of belief level $\alpha$. Formulated thusly, the rational recipient selects the trust level $\max_\beta U_R(\alpha, \beta)$ given $\alpha$. It follows trivially that in the limiting case, increasing the sanction level (thereby decreasing the supply of purveyors of false information) will have no effect on the trust level, and thus no effect on the harm generated by the expression of false speech. In case the recipient’s trust level is unaffected by a change in the proportion of purveyors of true and false information, neither the recipient’s harm nor the subject’s harm will change. Nevertheless, even when the extremum circumstance does not obtain, the potency effect will act as a friction, reducing the social benefit accrued from a reduction in the supply of false speech. Ergo, the potency effect can nullify the effect of changes to the relative composition of purveyors, and it will at least lessen the benefit of reducing the supply of defamers, however it cannot increase the total harm.

Next, even if the imposition of sanctions decreases the relative supply of purveyors of false information, the inverse supply effect will create a further friction on the social benefit of sanctions on defamatory speech. Assuming a prospective defamer’s extrinsic value is a function of the potency of his defamatory expressions, it follows that his incentives to defame will increase as recipients become more likely to believe his false statements. Thus, if sanctions are effective in affecting recipients’ trust levels, this will effect an increase in the incentives of defamers, creating a further friction on the effectiveness of sanctions.

Finally, even if the imposition of sanctions results in a net social benefit, despite the presence of the potency effect and inverse supply effect, there remains a further obstacle in filtering. Suppose there are two types of defamers: low-harm defamers and high-harm defamers. Let us define “low-harm” defamers as being those less

---

177 Assume $\alpha$ is the subjective probability that the signal is true.
credible than “high harm” defamers. In the face of sanctions, prospective low-harm defamers are more likely to refrain from engaging in defamatory speech than high-harm defamers, because they will receive less benefit from transmitting a defamatory signal. The imposition of sanctions will tend to dissuade more low-harm defamers from engaging in speech, and by the inverse supply effect, this will increase the incentives of high-harm defamers to exploit the increased credulousness of recipients. It follows that filtering will further reduce the effectiveness of legal sanctions, and in case the change in the composition of defamers (and the difference in magnitude between high-harm and low-harm) is sufficiently large, filtering may even result in a reduction in total social welfare.

Of course, if we assume away information costs, then the rational Bayesian recipient of information will strategically adjust his trust level to account for this as well, ensuring an equilibrium that represents a reduction in total social harm. However, I think the information costs associated with this are likely to be significant, given that the proportion of high-harm and low-harm defamers is likely to be opaque to prospective recipients. And I do not think this should be assumed out of the model. Nevertheless, even if we accept this assumption, although the net effect of filtering could not result in a reduction in social welfare, it would again in the extremum case render the imposition of sanctions null, and introduce yet another friction impairing the effectiveness of legal sanctions at a minimum.

Stepping back, the intuition here is easy to grasp. Prospective defamers crave a credulous audience. And the greater the proportion of defamers (relative to truth-tellers), the more skeptical the population of information recipients will be. Thus, allowing defamatory speech—or false statements of fact more generally—to go unregulated will tend to incentivize greater skepticism in the population of hearers, mitigating the harm to the recipients of potentially false information, and also mitigating the harm to subjects of potentially false information. It must be conceded that under some conditions, increased skepticism will reduce the benefit that recipients derive from true information (they will be skeptical of the true information insofar as they are unable to distinguish a true signal from a false signal without an independent search investment). Yet the point is not that the imposition of legal sanctions will have no effect (or a negative effect) on social welfare, but rather that plausible conditions exist, under which the imposition of legal sanctions could have no effect (or a negative effect) on social welfare.

And once again, even if the imposition of legal sanctions in some subset of defamation cases were welfare-improving, their effectiveness will tend to be undermined. And it is then plausible that the costliness of sanctions, litigation costs, and other kickers could tip the balance, such that the regulation of defamatory speech would
be net welfare-reducing when all factors are considered.

3.2.2 Commercial Speech

Expressive rights in the realm of commercial speech have historically enjoyed only attenuated protection. In particular, fraudulent inducements and deceptive advertisements are wholly outside the First Amendment guarantee. The argument for the received view is that false signals, designed to entice consumers to purchase a product, service, or property, may reduce incentives to enter into mutually beneficial exchanges. Unprotected from false signals, consumers will forgo many potentially Pareto-improving exchanges to avoid incurring the cost of information gathering required for verification of a signal’s truth. Furthermore, false signals can result in consequent harms arising from misplaced reliance on inferior or dangerous goods.

For example, if a tobacco company runs an advertisement claiming that cigarettes whiten teeth, freshen breath, and aid in children’s pulmonary development, then it could persuade some individuals to smoke. And if consumers later discovered that smoking actually increases the probability of developing various cancers, heart disease, emphysema, and rancid breath, then they would be reluctant to trust future promises of a product’s qualities and effects, refraining from entering into potentially Pareto-improving exchanges. When they do purchase goods, services, or property, they will at least invest more effort in search to validate the claims of advertisements.

Additionally, those individuals who are induced by the false advertisement to take up smoking might become addicted, continue the habit, and succumb to the maladies associated with cigarette smoking. The healthcare expenses which ensue are a further source of social cost, which regulations on deceptive advertising might prevent.

Restated somewhat more precisely, if the promised value of an exchange is $B$, the cost of disappointment is $C$, the investment in verification is $x$, and the probability that the signal is true is $p(x)$ such that $\frac{\partial p}{\partial x} > 0$, then consumers face an expected payoff of $p(x)B - (1 - p(x))C - x$ in the absence of regulation. Allowing disappointed consumers to collect damages when producers communicate false signals increases the consumer surplus. With regulation, consumers can expect a payoff of $B$.

The case for regulating fraudulent inducements and deceptive advertising is compelling. However, the cost in a deregulated regime may be somewhat less than the received argument predicts. The freedom to express false signals in a commercial context would not free promisors of their contractual obligations generally. It would simply be an abrogation of the promisee’s right to void a fraudulently induced agree-

\footnote{See supra \S 1.5.}
ment. In order to ensure liability for claims about a product, service, of property, the promisee—knowing he cannot rely on non-promissory claims—will simply insist that the claims be expressed in promissory terms.

Even still, there will be an increase in transaction costs and forgone surplus, and the harm-based reasons to favor regulation persist. This leads us to consider once again the possible countervailing effects. First, with respect to the potency effect, if false signals are regulated, then the ratio of true signals to false signals will tend to increase. Thus, consumers will decrease investments in verification and increase investments in reliance. This will magnify both consumers’ susceptibility to harm and the magnitude of harm when false signals are believed.

Second, with respect to the inverse supply effect, when unscrupulous sellers observe the increasing credulousness of consumers in the presence of regulation, they will want to exploit that gullibility. Additionally, for those products or services where the consumer’s reliance entails additional profit—for example, with subscriptions or brand ecosystems—sellers will have even greater incentives to engage in deceptive practices.

Third, with respect to filtering, it will tend to be the producers who sell the most inferior or dangerous goods who experience the strongest incentives to exploit the increased susceptibility of consumers in a regulated regime. Producers with quality products will simply refrain from overstating the excellence of their merchandise. Producers whose goods are only slightly substandard can more easily invest in improving quality or seek buyers who require products of lesser quality with honest advertising. The producer who has the most to gain from exploiting the greater credulousness of consumers is the producer whose goods are in least demand, and for whom it would be most costly to improve.

A full analysis of fraudulent inducement and deceptive advertising would require an excursus well exceeding the scope of this article. Additional complicating factors which a comprehensive investigation might include are: the effect of reputation, insurance, and private information screening services.

The case for regulation seems stronger for commercial speech than in the other speech contexts heretofore discussed. Yet even if the benefit of regulation were sufficient to overcome the countervailing effects, it is nevertheless important that lawmakers be cognizant of the tradeoffs in the design of commercial speech regulation. For example, devising a doctrine analogous to contributory negligence, such that consumers are obliged to undertake reasonable verification in order to claim damages, would help to mitigate underinvestment in search. And a “reasonable reliance” standard would help to mitigate excessive reliance investments.
3.2.3 Fake News

I should like to conclude my inquiry with a phenomenon which has arisen rather more recently: the problem of “fake news.” It is becoming ever more apparent that the Russian government in 2016 undertook an active disinformation campaign to bias the U.S. presidential election in favor of Donald Trump. Similar efforts have been observed in the “Brexit” referendum in the United Kingdom and in recent elections in the Ukraine, France, and Germany. There is little reason to suppose the Russian government will discontinue its activities, and it seems likely that foreign disinformation campaigns are likely to pose an ongoing threat to the democratic process. Indeed, it is a plausible surmise that the effectiveness of the Russian effort will inspire other states to embark upon similar escapades.

Much of the policy discussion concerning Russia’s disinformation campaign identifies it as a species of “cyber attack,” warranting defensive measures. There is of course an intuitive appeal to the proposition that the one must defend when attacked by hostile foreign actors. However, we should be wary of a knee-jerk response. Here again, attempts to combat persuasive speech will be likely to result in countervailing effects.

With respect to the potency effect, the analysis here is analogous to that in defamation and commercial speech. If measures aimed at reducing the supply of foreign disinformation are successful, then citizens will tend to invest less effort in search. Assuming that citizens want to acquire true information—or at least information not motivated by hostile motives—they will naturally tend to undertake greater investment in verification and less investment in reliance when there is a greater risk they are receiving foreign disinformation. Therefore, reducing the supply of foreign disinformation will tend to increase citizens’ vulnerability to the residual fake news which remains.

Next, with respect to the inverse supply effect, if the citizenry is made more credulous due to successful efforts to restrict the dissemination of foreign disinformation, then this will render it a yet more enticing target. Hostile foreign actors will be incentivized to redouble their efforts, engaging in more sophisticated and subtle mechanisms to influence voters, made more gullible by the reduction in the supply of fake news. Presumably the payoff from swinging an election would be substantial, and hostile foreign actors can be expected to invest considerable resources to exploit the impressionability of voters.

Finally, filtering may plausibly occur in at least two ways. First, relatively less hostile or less powerful foreign actors will be more likely to be deterred, leaving more hostile or more capable foreign actors to fill the void. A less hostile foreign actor will be more easily deterred, because its interests will be relatively more aligned with the
target nation’s interests. I take this to be what it means to be relatively less hostile. Those less hostile foreign actors will thus have less benefit from interfering in the target nation’s democratic processes. Less capable foreign actors will be deterred, because their efforts at disinformation, if not widely received, will tend to be less successful in generating the critical mass of mutually confirming counter-narratives which a comprehensive disinformation campaign would require.

Second, hostile foreign actors may strategically invest in disinformation only when the stakes are highest. Assuming efforts to curb foreign disinformation impose at least some cost upon the purveyors of disinformation, they will tend to focus their resources on efforts calculated to impose the greatest effect. They might save their efforts for those occasions when elections seem likely to be closely fought, or when a potential outcome is anticipated to be especially unfavorable.

It is of course possible that increasingly sophisticated disinformation campaigns may be countered with increasingly sophisticated regulation. However, this commits states to a rent-seeking game. Hostile foreign actors invest in more streamlined disinformation; and target states invest in more savvy regulation. The standoff is liable to lead to ever-escalating investments in expression and suppression, and a considerable dissipation of resources.

It is difficult to say with specificity how well strategies to combat foreign disinformation would fare in the face of the countervailing effects. Target governments are still developing their strategies. It is not yet clear whether such efforts would effect a reduction in supply even absent the countervailing effects. As governments work to formulate their responses to this fresh nuisance, I suggest they take seriously the option to do nothing. Governments should consider freely allowing hostile foreign actors to broadcast disinformation. Admittedly, the initial effect will tend to subvert the democratic process. However, a prevalence of foreign disinformation is, over time, likely to incentivize citizens to invest greater effort in verification, to reduce reliance, and generally to harden themselves against future subversive influences.

I suspect that the Russian government would rather deal with the regulation of social media than a population of skeptics.

4 Conclusion

This article reveals that the courts and doctrinal literature have embraced an implicitly economic framework in nearly every controversy concerning the freedom of expression. It is therefore a surprising irony that the economic analysis of speech rights has received so little attention from Law & Economics scholars. This article accepts the latent invitation, undertaking an explicitly economic exploration of the
speech right, extending the simplistic harm-deterrence model assumed by courts and prior scholarship to reveal a fuller, more capacious account of the issue.

The received view assumes that the speech realm is an implicit market. If the market analogy is taken seriously, the countervailing effects I identify go to the essence of the speech right and threaten to overwhelm the fundamental justification for speech regulation.

My argument does not entail an absolute prohibition on speech regulation. It merely undermines the logic of the received view. It is a theoretical point, raising the evidentiary burden for the proponent of regulation. More research is surely needed, and I hope some attempt is made to empirically verify or falsify the intuitive narratives I sketch in section 3. It may well be that some of the speech regulations I have discussed are (or would be) efficient law. The final determination whether a regulation is efficient requires empirical supplementation.

If we value the freedom of speech even one half as dearly as our hymns proclaim, it is incumbent upon the law that any diminishment of it be soundly justified. This article shows the defectiveness of the justifications upon which the law has hitherto relied. The mere harmfulness of an utterance cannot support incursions upon its expression.